bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Control of Hox transcription factor concentration and cell-to-cell variability by an auto-regulatory switch

Dimitrios K. Papadopoulos[1,*], Kassiani Skouloudaki[1], Ylva Engström[2], Lars Terenius[3], Rudolf Rigler[4,5], Vladana Vukojević[3] and Pavel Tomancak[1,*]

[1]Max-Planck Institute for Molecular Cell Biology and Genetics, 01307 Dresden, Germany

[2]Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, 10691 Stockholm, Sweden

[3]Center for Molecular Medicine (CMM), Department of Clinical Neuroscience, Karolinska Institutet, 17176 Stockholm, Sweden

[4]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17177 Stockholm, Sweden

[5]Laboratory of Biomedical Optics, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland

[*]Correspondence: papadopo@mpi-cbg.de and tomancak@mpi-cbg.de

Author contributions: DKP and VV, designed and performed the binding study of *Antp*, analyzed the data; DKP and KS, designed and performed the study of *Antp* relative transcript abundance, analyzed the data; YE, provided the *Antp* P1 reporter line; DKP and PT, conceived the comparative analysis of TFs and the examination of Antp variation in second instar imaginal discs; DKP, conceived the project, designed and performed all remaining experiments, analyzed the data, prepared figures, wrote the manuscript; DKP, KS, YE, LT, RR, VV and PT critically reviewed the manuscript and approved it for submission.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Concentration and DNA-binding kinetics of transcription factors (TFs) are crucial determinants of their function. Using Fluorescence Correlation Spectroscopy (FCS), we have characterized the concentration and cell-to-cell variability of 14 endogenously tagged TFs in live *Drosophila* imaginal discs. We found that variation in the concentration of different TFs among neighboring cells decreased with increasing average TF concentrations. Among the investigated TFs, Antennapedia (Antp), displayed the lowest average variability and transitioned from a low concentration/high variation state in second instar to a high concentration/low variation state in third instar imaginal discs. Using FCS and temporally resolved gain- and loss-of-function genetic studies we have uncovered that Antp is necessary and sufficient to drive a developmental regulatory switch from a state of auto-activation to a state of auto-repression. This mechanism is controlled by a progressive change in the relative concentrations of preferentially activating and repressing Antp isoforms that bind with different affinities to DNA.

Introduction

In order to understand the mechanisms that control pattern formation and cell fate specification in developing organisms, the intranuclear concentration, DNA-binding kinetics and cell-to-cell variability of relevant TFs need to be quantitatively characterized. TF concentration variability at the tissue level is thought to arise from diverse processes, including protein production and degradation. For example, gene transcription in a given tissue is generally a noisy process. The noise is both intrinsic to the gene expressed (Blake et al., 2003; Elowitz et al., 2002) and also due to variation arising from components of the post-transcriptional cellular machinery

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Swain et al., 2002). These can influence the expression and production of functional protein, resulting in protein concentration that exhibits variability among cells.

In undifferentiated tissue or cells, this TF cell-to-cell variability can be the driving force for differentiation. For example, progressive establishment of a Nanog salt-and-pepper expression pattern leads to the formation of primitive endoderm in the mouse preimplantation embryo, whereas loss of the variation results in embryos lacking primitive endoderm entirely (Kang et al., 2013). In Drosophila, *Senseless* (*Sens*) expression is required for the establishment of proper number of sensory organ precursors in the ectodermal proneural cluster of cells and unequal concentration among cells is required for their specification (Li et al., 2006). Finally, variation in concentration (rather than its overall average concentration) of the Yan TF causes the transition of developing photoreceptor cells to a differentiated state during Drosophila eye development (Pelaez et al., 2015).

Conversely, in already differentiated tissue or cells, which have acquired a steady state during development, TF variability may need to be counteracted across cells to ensure homogeneity of gene expression patterns and robustness of commitment to a certain transcriptional regime. Such a function has been identified for the Snail (Sna) TF, which is required for the invagination of the mesoderm during Drosophila gastrulation (Boettiger and Levine, 2013), or the Bicoid (Bcd) and Hunchback (Hb) TFs during early embryogenesis (Gregor et al., 2007a; Gregor et al., 2007b; Little et al., 2013). These studies have quantified the tolerable degrees of concentration variability allowing establishment of gene expression territories of remarkable precision in the developing embryo.

Finally, differential fates within the same developmental territory may be induced by TFs deploying different DNA-binding dynamics in between seemingly

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

equipotent cells, despite the existence of very similar concentrations (i.e. low variability). For example, studies on the Oct4 TF in early mouse embryos have shown that differential kinetic behavior of DNA binding, despite equal Oct4 concentration among blastomeres, ultimately dictates an early developmental bias towards lineage segregation (Kaur et al., 2013; Plachta et al., 2011).

So far, variation studies have focused predominantly on monitoring the noise of mRNA production and have dealt almost exclusively with early embryonic patterning TFs and morphogen gradients (Holloway et al., 2011; Holloway and Spirov, 2015; Little et al., 2013; Lucas et al., 2013; Pare et al., 2009). Little information exists about TF variability at the protein level within a tissue, since studies of this sort would require single-molecule sensitivity. Even more importantly, proteins need to be quantified at endogenous levels (Lo et al., 2015).

High sensitivity methods for quantification of proteins of low intracellular abundance, like TFs (Kim et al., 2009; Vaquerizas et al., 2009) usually rely on proteomic analyses based on mass spectrometry. Despite the validity of these methodologies and their evolution to acquire higher accuracy (Brun et al., 2007) and resolution (Hanke et al., 2008), they are inappropriate for studies of variability among cells in the same tissue. They additionally cannot reach the level of single-molecule detection and, most importantly, they cannot be easily applied in developing tissues or cells.

We have previously used methods with single-molecule sensitivity, Fluorescence Microscopy Imaging and FCS, to study Hox TF interactions with nuclear DNA in living salivary gland cells (Papadopoulos et al., 2015; Vukojevic et al., 2010). FCS has been instrumental for quantification of TF dynamics in living cells or tissue in several recent studies (Clark et al., 2016; Kaur et al., 2013; Lam et al.,

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2012; Mistri et al., 2015; Papadopoulos et al., 2015; Perez-Camps et al., 2016; Szaloki et al., 2015; Tiwari et al., 2013; Tsutsumi et al., 2016). Yet, in these studies, measured protein mobility in overexpressing systems, rather than molecular numbers at endogenous levels.

Recent progress in Drosophila genome engineering led to the establishment of genome-scale resources of endogenously-tagged proteins, modified at the endogenous loci, or by addition of a tagged third copy allele within large genomic fragments (Buszczak et al., 2007; Ejsmont et al., 2011; Ejsmont et al., 2009; Kelso et al., 2004; Morin et al., 2001; Quinones-Coello et al., 2007; Sarov et al., 2016; Venken et al., 2011). These transgenic collections enable, for the first time, systematic studies of TF protein dynamics *in vivo* at endogenous levels, using methods with single-molecule sensitivity.

In this study, we take advantage of the availability of fly toolkits, in which TFs have been endogenously tagged by different methodologies (6 fosmid lines, 6 BAC lines, 1 FlyTrap line and 1 MiMIC line), to measure the intranuclear concentration of various TFs in vivo by FCS, and their cell-to-cell variability in several fly imaginal discs. The imaginal discs are flat single-layer epithelia comprised of small diploid cells and many TFs are expressed in defined regions within these tissues during development. These TFs were selected on the basis of the following criteria: a) availability of an endogenously-tagged line, b) well characterized expression patterns in imaginal discs, c) representation of most major imaginal discs in our dataset (wing, leg, eye, antennal, genital discs, and also salivary glands) and d) ability to readily visualize them using a standard confocal microscopy setup. Our study identifies Antp as a TF displaying notably lower variability among neighboring cells at the level of

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

protein concentration compared to other TFs and uncovers a mechanism that controls Antp concentration and variation in developing discs.

Results

Characterization of average protein concentrations and cell-to-cell variability of 14 Drosophila TFs

Average concentrations of TFs in neighboring nuclei of third instar imaginal discs were measured by FCS (Figure 1). FCS measurements were preferentially performed in the domains of highest expression in the discs (Figure 1 A-P, white arrows) and have sampled adjacent cell nuclei within these domains. Temporal autocorrelation curves differed among TFs in their amplitude (Figure 1Q), characteristic decay times and the relative contribution of the slow component (Figure 1R). They were fitted with a two-component model of free three-dimensional diffusion and triplet formation (Supplement 1) and interpreted as described in detail in our previous study (Vukojevic et al., 2010). Differences in average concentrations ranging about two orders of magnitude among different TFs, from $\sim 30\ nM$ to $\sim 1.1\ \mu M$ (see Supplement 2 for the calculation) were observed.

FCS analysis of inter-nuclear variability in TF concentration among adjacent nuclei (Figure 2) showed that variation, expressed as variance over squared mean ($CV^2 = \frac{s^2}{m^2}$), ranges from ~7 to ~53% (Figure 2 A). We also examined the protein Fano factor values of this dataset (expressed as variance over the mean, $F_f = \frac{s^2}{m}$) and obtained values ranging from ~0 to ~20 (Figure 2 A'), in line with previous observations (Sanchez et al., 2011). To evaluate a possible relation of variation to

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

concentration among different TFs, we plotted the variation as a function of average TF concentration (Figure 2 B). We observed elevated levels of variation ($CV^2 > 0.2$) for most TFs, and the variation decreased with increasing average concentrations for all TFs (Newman et al., 2006; Schwanhausser et al., 2011; Taniguchi et al., 2010). But, we also noted that Antp displayed lower variation values for its measured average concentration ($CV^2 < 0.2$), as compared to the other TFs in our dataset (Figure 2 A and B). This distinction prompted us to further examine variation values for average endogenous Antp concentrations among neighboring cells in a collection of cell clusters in the wing and leg discs (Figure 2 C). We observed that the variation decreased steeply with increasing Antp concentration (Figure 2 C), whereas the protein Fano factor increased (Figure 2 C'). This behavior, which is usually indicative of complex regulatory processes at the transcriptional level (Franz et al., 2011; Smolander et al., 2011), led us to further investigate the regulation of Antp transcription genetically.

Control of Antp concentration by auto-repression

We started investigating possible regulatory mechanisms of Antp abundance in imaginal discs by examining whether Antp is able to control its own levels (auto-regulation). To do this, we first established that the Antp-eGFP MiMIC line, used in our FCS measurements, represents a functional *Antp* allele. This fly strain has been generated by conversion of the MiMIC insertion MI02272 to a protein trap insert (Venken et al., 2011). Since critical domains for Antp function, such as the homeodomain and YPWM motif, are not interrupted by the introduction of the eGFP artificial exon (Figure 3 – figure supplement 1), we reasoned that the protein should retain a great portion of its full functionality *in vivo*. This is indeed the case for

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

178 synthetic Antp (SynthAntp) peptides lacking the long non-conserved N terminus of
179 the protein, which we have functionally investigated previously (Papadopoulos et al.,
180 2011). In fact, flies bearing the MI02272-eGFP protein trap are homozygous viable
181 with adult escapers and only partial lethality at the pupal stage. Because *Antp* mRNA
182 is not maternally deposited, this indicates that at least the embryonic, larval and
183 pupal functions of *Antp* are retained in the Antp-eGFP fusion construct. To
184 investigate whether the Antp-eGFP pattern truly reflects the expected Antp protein
185 distribution, we performed immunostainings of the embryonic ventral nervous system
186 and all thoracic imaginal discs (pro-, meso- and metathoracic leg, wing, haltere and
187 humeral discs) of wild type and heterozygous Antp-eGFP larvae. We observed that
188 the Antp-eGFP fusion protein recapitulates the endogenous *Antp* expression pattern
189 (prominent colocalization of eGFP and Antp, Figure 3 – figure supplement 2). The
190 expression patterns are similar to the ones observed for *Antp* mRNA, reporter
191 constructs and protein (Engstrom et al., 1992; Levine et al., 1983; Wirz et al., 1986)
192 in third instar imaginal discs. Together, this data support the notion that Antp function
193 is unaltered in Antp-eGFP MiMIC flies.

194 In order to address whether Antp is able to auto-regulate its own protein
195 levels, we monitored the concentration of endogenous Antp protein upon
196 overexpression of *Antp*. To distinguish between overexpressed and endogenous
197 protein, we used SynthAntp transgenes fused to eGFP (SynthAntp-eGFP). These
198 transgenes encode the region between the YPWM motif of Antp and its C terminus
199 (but lack the long and non-conserved N terminus of the protein, against which Antp
200 antibodies have been raised). Importantly, their Antp-specific homeotic function has
201 been validated previously (Papadopoulos et al., 2011). Clonal overexpression of

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SynthAntp-eGFP in the wing disc (Figure 3A-C') repressed the endogenous Antp protein, indicating that Antp is indeed able to regulate its own protein levels.

Antp protein can repress *Antp* gene expression at the transcriptional level

We then asked whether this repression occurs at the transcriptional level. To address this we used a previously characterized reporter of the *Antp* P1 promoter, *Antp* P1-*lacZ* (Zink et al., 1991), containing 17.2 kb of the P1 promoter and the first intron of *Antp* (see Materials and Methods). This reporter has been shown to recapitulate the expression pattern of the *Antp* mRNA from the P1 promoter ((Engstrom et al., 1992; Zink et al., 1991) and unpublished results). We induced overexpression clones of SynthAntp-eGFP in the wing disc and observed that *Antp lacZ*-reporter is repressed in the clones (Figure 3D-F). The auto-repressive capacity of Antp was also validated for all leg discs where *Antp* is strongly expressed (pro-, meso-, and metathoracic leg discs), using SynthAntp-eGFP and three other Gal4 drivers, *Distal-less*- (*Dll*), *patched*- (*ptc*) and *69B*-Gal4 (Figure 3 – figure supplement 3).

In order to rule out that the observed repression is due to putative dominant negative activity of SynthAntp-eGFP, we repeated the experiment in overexpression clones of the full-length untagged Antp protein that is functionally equivalent to the endogenous Antp. Nevertheless, similarly to SynthAntp-eGFP, we observed repression of the *Antp* reporter (Figure 3G-I'), suggesting that the transcriptional auto-repressive capacity of Antp is not caused by dominant negative effects of the minimal *Antp* transgene.

Without excluding additional post-transcriptional mechanisms of auto-repression and with no major Antp auto-regulatory capacity having been identified

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

using a P2 promoter reporter (data not shown) (Engstrom et al., 1992; Zink et al., 1991), our results provide evidence that Antp is able to repress itself from the P1 promoter at the transcriptional level.

Antp protein is able to activate *Antp* gene transcription in imaginal discs

Antp auto-repression capacity offered a plausible explanation for the observed low variation in Antp protein concentration among third instar cell nuclei. However, overexpression of *SynthAntp-eGFP* from the *Distal-less* (*Dll*) (MD23) enhancer surprisingly resulted in ectopic activation of the *Antp* P1 reporter in distal compartments of third instar discs, where Antp is normally not detected, such as the wing pouch and the distal portion of the antennal disc (Figure 3 J-M, O-R and eGFP controls in N, S and T-V). This finding is opposite to the aforementioned auto-repressive function of Antp and suggested a role in transcriptional auto-activation.

Importantly, the MD23 *Dll*-Gal4 driver used is not a homozygous viable line and the Dll TF is itself a repressor of *Antp* during disc development (Emerald and Cohen, 2004). Therefore, to preclude that the observed P1 activation is due to a Dll sensitized background, we induced *Dll RNAi*-knockdown clones and confirmed that the absence of Dll is not sufficient to induce *Antp* ectopic expression or alter its capacity to repress in the endogenous expression domain (Figure 3 – figure supplement 4).

We conclude that, in addition to its capacity to repress, Antp is also able to activate its own transcription.

A temporal switch controls the transition of *Antp* from a state of auto-activation to a state of auto-repression

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We reasoned that, since endogenous Antp were able to control its own expression both positively and negatively, this is likely to occur in a temporally resolved manner. To investigate whether auto-regulation of Antp is temporally controlled in development, we induced overexpression clones of full-length untagged *Antp* either at 26 h (early – first larval instar) or at 60 h (late – second larval instar) of development and harvested third instar wandering stage larvae for analysis (Figure 4 A). Endogenous Antp protein was monitored by means of eGFP visualization of the MiMIC allele. Induction of clonal overexpression of untagged *Antp* in discs early in development, just after embryonic hatching (first instar), displayed strong ectopic activation of Antp-eGFP (Figure 4 B-D). Notably, in the case of early-induced clones with auto-activation readout, we found no indication of concurrent repression of Antp within its normal expression domain (dashed red line in Figure 4 E and G). Rather, we observed only Antp activation throughout the disc, even in the *Antp* normal expression domain. We also confirmed that early auto-activation of Antp takes place at the transcriptional level (Figure 4 – figure supplement 1). In contrast, overexpression clones induced during the late second instar stage repressed Antp-eGFP (Figure 4 H-J).

Together, these results clearly pointed to the existence of a switch in Antp auto-regulatory capacity on its own transcription during development. Starting from a preferentially auto-activating state, it changes into an auto-inhibitory one, at later developmental stages. They also suggested that there is no major temporal overlap between auto-activation and auto-repression.

<u>Antp is required for activation and repression of *Antp* gene transcription during development</u>

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Clonal overexpression experiments provided useful insight as to the ability (i.e. the sufficiency) of Antp protein to display opposing regulatory effects on its own transcription during development. We next sought to confirm that the changes in *Antp* expression actually require Antp protein, and thereby validate the auto-regulatory loop. To this end, we first established that *Antp* can be efficiently knocked down clonally in the wing disc by monitoring the repression of the MiMIC-derived Antp-eGFP protein (Figure 4 K-L'). Then, we induced early *Antp RNAi*-knockdown clones, monitored the activity of the *Antp* P1 reporter and found its expression to be efficiently decreased (Figure 4 M-N' and eGFP controls in Q-R'). This result is in agreement with the results from early clonal overexpression of the full-length protein, which rather triggers activation of *Antp* P1 transcription (Figure 4 B-G and Figure 4 – figure supplement 1). We conclude that Antp is required for sustaining its own expression, confirming the existence of an *Antp* auto-regulatory loop.

Conversely, we found that, when *Antp* was clonally down-regulated by *RNAi* at later developmental time points (Figure 4 A), the transcription of the reporter was rather up-regulated in the knockdown clones (Figure 4 O and P). Again, this result is in agreement with the corresponding over-expression experiments, which resulted in repression of Antp protein and transcripts (Figure 3 A-I', 4 B-D and Figure 3 – figure supplement 3). Hence, next to Antp being sufficient, we have shown that it is also required for both repression and activation of *Antp* P1 transcription in thoracic imaginal discs. Thus, normal patterns of *Antp* expression require both auto-activation and auto-repression.

<u>Direct correlation between Antp concentration and homeotic function</u>

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

With Antp having been identified to possess opposing roles on the regulation of its transcription and the observation of prominent auto-activation in some overexpression assays (Figure 3 J-V and Figure 4 – figure supplement 1), but no obvious auto-activation in others (Figure 3 – figure supplement 3), we reasoned that these differences might be due to a) changes in concentration (i.e. auto-repression and auto-activation might occur at different concentrations of Antp) and b) differences in DNA-binding affinities associated with activation and repression.

To examine the effect of concentration on Antp function, we expressed the same *SynthAntp-eGFP* construct in the antennal disc, using four different Gal4 driver lines, *Dll* (MD23)-, *ptc*-, *Dll*- (MD713)- and *69B*-Gal4. We monitored the concentration of SynthAntp-eGFP by imaging (Figure 5 A-D) and FCS (Figure 5 I and I'), while also assessing its capacity to induce homeotic transformations in the antenna (Figure 5 E-H). We found that the *Dll* (MD23) line is indeed the strongest driver among the four (Figure 5 A). It caused complete transformations of the distal antennal segments into a tarsus (Figure 5 E), whereas the remaining Gal4 drivers all induced transformations to different extents, progressively weaker for decreasing concentrations (Figure 5 B-D and F-H). The concentration range between the strongest and weakest Gal4 driver was more than eightfold (Figure 5 A-D, I and I'). FCS analysis also showed that, within the experimental error of our measurements, there were no differences among Antp FCS measurements upon induction by *ptc*-Gal4, *Dll*-Gal4 and *69B*-Gal4 (Figure 5 J), as evident from the overlap of the temporal autocorrelation curves shown in Figure 5 I after their normalization to the same amplitude, $G_n(\tau) = 1$ at $\tau = 10\ \mu s$ (Figure 5 J). This indicated that Antp DNA-binding behavior is not altered by overexpression.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Taken together, these results show that different Gal4 drivers yield highly different concentrations of overexpressed Antp (Figure 5 I and I'). Moreover, the homeotic activity of Antp readily depends on, and directly correlates with, the concentration triggered by ectopic activation (Figure 5 E-H).

Antp can repress and activate its transcription at endogenous concentrations

If Antp auto-activation is readily observed only in conditions of very high overexpression ($3.8\ \mu M$), more than 7 times higher than the average endogenous concentration of Antp-eGFP in the wing disc measured by FCS ($0.5\ \mu M$, Figure 1 K and Q, light grey autocorrelation curves), it is possible that Antp auto-repression and auto-activation do not occur at all at physiological Antp concentrations. To analyze whether this is the case, we carefully examined the capacity of Antp to repress and ectopically activate its own expression using the *69B*-Gal4 driver. This line triggered concentrations of overexpressed Antp of $473\ nM$ in average (Figure 5 D), similar to endogenous Antp levels ($501\ nM$, Figure 1 K and Q). We confirmed that in the wing disc, endogenous Antp protein is repressed at the base of the wing blade, when it is co-expressed with the *69B*-Gal4-driven SynthAntp (Figure 5 K-L, negative controls in M-N and Figure 3 – figure supplement 3). However, we did not observe any ectopic *Antp* P1 reporter activation caused by *69B*-Gal4-mediated overexpression using antibody staining against β-galactosidase (Figure 3 – figure supplement 3 and Figure 3 – figure supplement 4, E-F'). By performing more sensitive X-gal stainings of *Antp* P1-lacZ activity however, we detected weak but definite induction of *Antp* P1 expression in the antennal disc (Figure 5 O and negative control in P). We also confirmed the ability of the remaining two Gal4 drivers (the MD713 *Dll*-Gal4 and *ptc*-Gal4) to activate *Antp* transcription ectopically (Figure 5 – figure supplement 1). We

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

conclude that Antp is able to repress and activate its own transcription at physiologically relevant concentrations, comparable to its normal concentration in the wing disc. Moreover, our data suggest that, for equal Antp concentrations, auto-repression is a more efficient process than auto-activation.

Antp switches from a low-concentration/high-variation to a high-concentration/low-variation state

The *Antp* gene is able to switch between states of auto-activation and auto-repression and it is likely that the auto-repressive state limits the variation of Antp protein concentration among neighboring cells late in development, in third instar imaginal discs. If so, variation would be higher during earlier stages, when auto-repression does not operate. To test this hypothesis, we monitored the endogenous expression levels and cell-to-cell variation of Antp nuclear concentration in second instar wing and leg discs by FCS, using the Antp-eGFP MiMIC allele (Figure 6 A-D). We observed significantly lower average concentrations of Antp protein in both second instar wing and leg discs, as compared to their third instar counterparts (Figure 6 C and D, compare the amplitudes of the average autocorrelation curves – the amplitude is inversely proportional to the number of molecules). This indicates that endogenous Antp concentration changes from second to third instar in thoracic discs (Figure 6 C). Importantly, variation in concentration among neighboring nuclei in second instar discs was roughly double compared to third instar discs (Figure 6 C), indicating that the increase in concentration is accompanied by a suppression of concentration variability.

FCS measurements were performed either on endogenous Antp in second and third instar wing and leg discs (Figure 6 D and E); or on endogenous and

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

overexpressed Antp in the wing and antennal discs, respectively (Figure 6 F and G). Temporal autocorrelation curves recorded in third instar discs, when normalized to the same amplitude, $G_n(\tau) = 1$ at $\tau = 10\,\mu s$ (Figure 6 E), displayed a notable shift towards longer characteristic times, as compared to measurements performed in second instar discs. Fitting of the autocorrelation curves (see Supplement 1 for details) identified two average characteristic decay times, $\tau_1 = (800 \pm 400)\,\mu s$ and $\tau_2^{2nd\ instar} = (100 \pm 150)\,ms$, whereas in the third instar average τ_2 was found to be significantly longer, $\tau_2^{3rd\ instar} = (250 \pm 250)\,ms$ (Student's two-tailed t-test, $p = 3 \cdot 10^{-7}$). This observation indicates that endogenous Antp is initially fast moving in the nucleus and participates in considerably fewer interactions with the DNA. However, in later discs, it appears to display much longer decay times, reflected by the shift of τ_2 to longer times, indicating longer-lasting interactions with the DNA. In contrast, when FCS measurements were compared between endogenous third instar and overexpressed Antp (Figure 6 F and normalized autocorrelation curves in G) we found indistinguishable decay times, as evidenced from the overlap of the normalized temporal autocorrelation curves, despite more than sevenfold concentration difference ($0.5\,\mu M$, as compared to $3.8\,\mu M$). Fitting of the autocorrelation curves (see Supplement 1 for details) identified average $\tau_2^{o/e} = (300 \pm 300)\,ms$, similar to the average value obtained from fitting third instar FCS data (Student's two-tailed t-test, $p = 3.3 \cdot 10^{-1}$). Moreover, the characteristic decay times were found to be independent of the concentration of Antp-eGFP over the whole concentration range examined (Figure 6 – figure supplement 1, A and B), showing that the underlying fluorescence intensity fluctuations are generated by diffusion.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our results show that *Antp* starts being expressed at relatively low and highly variable levels in early developing discs, and with the capacity to auto-activate its own transcription. Later in development, it reaches a state of higher average concentrations and considerably lower variation between neighboring cells.

Calculation of the endogenous apparent dissociation constant of Antp from FCS measurements

With concentration of Antp not being sufficient to explain the observed differences in Antp dynamics in the second and the third instar leg and wing discs (compare Figure 6 E with G), we sought to investigate the underlying differences at the level of Antp-DNA binding.

Live imaging of Antp-eGFP in the wing disc (Figure 6 H and I) revealed that Antp is not evenly localized in cell nuclei, it is excluded from the nucleoli and forms bright sites of accumulation in different nuclear compartments (Figure 6 K and L). These are presumably sites of pronounced DNA-binding and transcriptional regulation. Taking advantage of the variation in Antp-eGFP concentration among wing disc nuclei, we constructed DNA-binding curves and calculated the apparent endogenous dissociation constant for Antp in third instar discs (Figure 6 M). We observed that the fraction of DNA-bound Antp, which is reflected by the relative amplitude of the slow component measured by FCS, scales linearly with the total amount of TF (Figure 6 M and Figure 6 – figure supplement 1, A), yielding the following regression equation: $y = 0.38x - 5.78$. Using the previously described simple two-step model of TF-DNA binding, in which non-specific interactions are faster and precede the specific interactions (Vukojevic et al., 2010), we have calculated the endogenous apparent dissociation constant for specific third instar

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Antp binding to nuclear DNA to be $K_{d,Antp}^{3rd\ instar} = 18\,nM$, (see Supplement 3 for the calculation of the specific dissociation constant and Supplement 4 for the approximation of the number and concentration of specific binding sites for Antp). Importantly, we arrived at a very similar value of the apparent dissociation constant, when it was calculated from the sum of FCS measurements on third instar endogenous and overexpressed Antp (Figure 6 H-J, 209 measurements), namely $K_{d,Antp}^{endog.+o/e} = 15\,nM$ (Figure 6 N and Figure 6 – figure supplement 2, B, regression equation: $y = 0.46x - 7.29$, see Supplement 5 for the calculation). Fluorescence Recovery After Photobleaching (FRAP) analysis (Figure 6 – figure supplement 1, C) showed that the dissociation rate of specific Antp-DNA complexes is low ($k_{off}^{specific} < 3 \cdot 10^{-3}\,s^{-1}$). Together, FCS and FRAP measurements indicate that Antp binds with high affinity to its specific binding sites in third instar disc cells (Figure 6 – figure supplement 1, C), as evidenced from the apparent dissociation constant that is in the order of $10\,nM$. In contrast, the dissociation constant calculated from FCS measurements performed in second instar discs was $K_{d,Antp}^{2nd\ instar} = 131\,nM$ (Figure 6 N and Figure 6 – figure supplement 2, C, regression equation: $y = 0.44x - 2.6$, see Supplement 6 for the calculation), indicating considerably lower affinity of Antp binding to the DNA in early discs.

We conclude that Antp binding behavior to the DNA changes from second to third instar discs, transitioning from a state of low affinity to a state of stronger binding on the DNA.

<u>Dynamic control of Antp auto-regulation by different Antp isoforms</u>

The changing binding behavior of Antp on DNA from second to third instar discs and the developmental transition from an auto-activating to an auto-repressing

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

state suggested a causal relationship between the two parameters. However, it provided no explanation as to how this could be regulated molecularly. We, therefore, sought to identify additional regulatory mechanisms that could link the observed differences in DNA-binding affinities to Antp auto-activation and repression.

We have previously established that the transcriptional activation and repression capacities of the Antp TF are favored or disfavored by naturally occurring isoforms differing in as little as four amino acids in the linker between the YPWM motif and the homeodomain (Papadopoulos et al., 2011). Long linker isoforms (8 amino acids) favor transcriptional activation of Antp target genes, whereas short linker ones (4 amino acids) favor repression (Figure 7 A). We, therefore, examined whether the linker length is also responsible for differences in auto-regulation. Ectopic expression of SynthAntp-eGFP peptides featuring a long linker displayed significantly weaker repression capacity on endogenous Antp, as compared to their short linker counterparts (Figure 7 B-E' and quantified in J). This finding indicates that, also in the case of auto-repression, short linker isoforms function as more potent repressors. Interestingly, the YPWM motif was found to be to a great extent dispensable for auto-repression. Constructs bearing mutations of the YPWM motif to alanines (Figure 7 A) could still repress endogenous Antp (Figure 7 F-I' and quantified in J). We conclude that different Antp isoforms display differential auto-repressive capacity and that this function resides in a small C-terminal portion of the protein, which includes the homeodomain, whereas the long N terminus and the YPWM motif are dispensable.

If Antp auto-repression is differently regulated by short and long linker isoforms, this may be reflected in the relative abundance of their corresponding

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mRNA variants. We, therefore, investigated the expression of long and short variants in thoracic imaginal discs by quantifying their relative abundance. We found that both transcript variants are expressed from embryonic to third instar development (Figure 7 K). However, a major redistribution of their relative amounts occurs during this time, starting with excess of long linker transcripts and ending with excess of short linker ones in third instar wing and leg imaginal discs, in line with previous observations (Stroeher et al., 1988).

With short linker isoforms showing enhanced repressor capacity and long linker ones showing enhanced activating capacity, our findings suggest that the switch between auto-activation and auto-repression is controlled at the level of transcript variant abundance (without excluding additional mechanisms).

Relative changes in Antp transcript variant concentration, differential efficiency of their encoding isoforms to repress or activate the *Antp* gene, the developmental switch of the *Antp* gene from auto-activation to repression and the different mobilities of Antp between second and third instar imaginal discs all point towards the hypothesis that the two isoforms have different properties in their modes of interaction with the DNA. To investigate this hypothesis, we expressed the two isoforms in third instar wing and antennal discs from the *69B* enhancer, which we established to result in Antp concentrations close to (if not below) endogenous levels (Figure 5 D). We performed FCS measurements, sampling cells with similar expression levels of short or long linker *Antp-eGFP* isoforms (Figure 7 I'). We observed that the short linker isoform displayed shorter characteristic decay times, as evidenced by the shift of average temporal autocorrelation curves towards longer characteristic decay times and the higher relative contribution of the second component as evident from its increased relative amplitude (Figure 7 L'). Similar to

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

our comparative analysis between second and third instar discs (Figure 6 D and E), this finding indicates that the short linker isoform binds to the DNA with considerably higher affinity than its long linker counterpart. Plotting the fraction of the slow FCS component (reflecting the DNA-bound fraction of Antp), as a function of the total Antp concentration (as in Figure 6 M-O) allowed us to calculate the apparent equilibrium dissociation constants for the two isoforms (see Supplement 7 for the calculation) to be $K_{d,Antp}^{short\ linker\ isof.,o/e} = 18nM$ (Figure 7 M and Figure 7 – figure supplement 1, D, regression equation: $y = 0.34x - 5.31$) and $K_{d,Antp}^{long\ linker\ isof.,o/e} = 190nM$ (Figure 7 N and Figure 7 – figure supplement 1, E, regression equation: $y = 0.24x - 3.28$). The tenfold difference in the dissociation constants indicates a tenfold higher binding affinity of the short linker isoform on the DNA, in line with our observations during the developmental transition from second to third instar discs (Figure 6 M-O).

Taken together, our experiments on endogenous and synthetic Antp and on short and long linker variants indicate that differences during disc development can be largely attributed to differences in the affinity of the investigated isoforms.

We have also addressed the efficiency of short and long linker variants in Antp auto-activation during early stages (in the embryo) (Figure 7 – figure supplement 2) and found that the long linker isoform displayed stronger activation capacity of the *Antp* P1 reporter than its short linker equivalent, but no ectopic activation was observed upon mutation of the YPWM motif (Figure 7 – figure supplement 2).

Taken together, the switch of Antp from an auto-activating to an auto-repressive state and the alteration of its DNA-binding behavior during disc development can be largely explained by a developmental regulatory shift in the

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

relative concentrations of preferentially activating and repressing Antp isoforms, which themselves display distinct properties in their modes of interaction with the DNA.

Discussion

In this work, we have characterized the endogenous molecular numbers (concentration) and cell-to-cell variation in the concentration of 14 *Drosophila* TFs in imaginal discs. With single-point FCS being routinely used for measurements of the concentration of fluorescent molecules and their dynamics in live cells, and with novel FCS methodology for larger scale measurements across a whole tissue or animal constantly emerging (Krieger et al., 2015; Papadopoulos et al., 2015), we expect that the study of variation in protein levels will become readily accessible, complementing the established methods to quantify mRNA variation.

We focused our in-depth mechanistic study on Antp and our findings are summarized in Figure 8. *Antp* is initially expressed in thoracic imaginal discs, producing (at least) two functionally distinct isoforms, one being a preferentially auto-activating and the other a preferentially auto-repressing species. The two isoforms differ in the mode of interactions they undergo with the DNA, as well as their affinities of DNA-binding, with the activating isoform characterized by low binding affinity and the repressing isoform showing higher binding affinity on the DNA. The differences in affinities between the two isoforms could be owed to a) *trans* modifications (one isoform being a more efficient DNA-binder than the other), or b) to *cis* modifications of the corresponding DNA-binding sites at the chromatin level (binding sites being more receptive to binding by one Antp isoform and not the other). Moreover, the two isoforms appear to be initially present in the nuclei in unequal concentrations (higher

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

concentration of the activating isoform). For simplicity, the binding of Antp isoforms to the Antp promoter is depicted as direct, although an indirect mechanism, involving induction or repression of intermediate genes, is also possible. Moreover, binding sites of activating or repressing isoforms may or may not be the same. The aforementioned state is characterized by overall low average concentrations of Antp and high variability of total concentration among neighboring cells. High variation in endogenous Antp protein concentration suggests a high level of transcriptional noise introduced in the regulation of Antp target genes. In fact, it has been well established that increased affinity of TF binding to its target sites readily increases transcriptional noise (Dadiani et al., 2013; Shimoga et al., 2013; Suter et al., 2011).

With *Antp* transcription being induced in the majority of the cells (without ruling out repression in some of them), the total amount of TF increases and (by means of an unknown regulatory mechanism) the relative distribution of preferentially activating and repressing Antp species is progressively shifted towards the production of more repressing isoforms. Therefore, above a certain concentration threshold, the Antp promoter gradually switches from an activated configuration to a repressed one, as development proceeds. Because this switch is concentration-dependent, it results in higher homogeneity of Antp concentration among cells. Therefore, variation becomes progressively suppressed.

The gradual repression of *Antp* at the transcriptional level provides a plausible explanation for the measured high variation in concentration at early, and the low variation at later, stages. At the same time, our measurements suggest that no steady state in Antp concentration has been reached until, at least, early third instar development. With the repressor isoforms displaying stronger and longer-lasting

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

binding to the DNA, this state persists, allowing the control of variation, while maintaining Antp at desired levels.

Taken together, our quantitative measurements of protein concentrations, cell-to-cell variation, DNA-binding affinities and differential regulatory capacities of isoforms suggest a simple qualitative model of auto-regulation in the *Antp* locus. This model should explain how Antp transits from an initial state of low concentration/high variation to a state of high concentration/low variation. Future experiments, in combination with quantitative stochastic modeling, will test whether the proposed model is indeed sufficient to explain the establishment of a stable, high concentration/low variation state of a TF such as Antp.

Is Antp maintenance required for proper development of thoracic discs? In our overexpression experiments in leg and wing discs, we showed that increased Antp concentration during disc development leads to loss of Antp transcripts and protein, which is normally expressed in defined domains throughout disc development. Despite this, flies strongly overexpressing SynthAntp by Dll-Gal4 (MD23) are viable and fertile and can be kept as a stock. The lack of adult Antp loss-of-function phenotypes (due to repression of endogenous *Antp*) in flies strongly overexpressing SynthAntp requires explanation. One possibility is that Antp function is dispensable during late disc development. A second possibility is that, despite repression of the endogenous *Antp* gene, *SynthAntp* rescues its function. We addressed this question and found that SynthAntp can indeed compensate to a great extent for repression of endogenous Antp (Figure 8 – figure supplement 1). Our results indicate that Antp maintenance is required for normal leg and notum development and suggest that endogenous auto-repression within the *Antp* normal expression domains is in fact an

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

inefficient process, which takes place upon concentration increase only (for example, as a cell-autonomous response to increased variation).

In our dataset of 14 TFs, we have measured values of variation in TF concentration ranging from 40% to 60% in imaginal discs and determined their corresponding protein Fano factor values. Having measured variation of molecular numbers, the obtained data represent true Fano factor values, devoid of scaling factors (Raj and van Oudenaarden, 2009; Sanchez et al., 2011). Our study shows that Fano factors values for all TFs measured are in the range of 0-20, in line with Fano factor values of other TFs determined previously to lie between 0 and 30 (Sanchez et al., 2011). Moreover, the majority of TFs examined show high Fano factor values, $F_f > 1$, providing strong evidence that transcriptional bursting is likely to be a significant source of the observed cell-to-cell variability.

Contrary to Sens, Fkh and Antp TFs, which displayed $F_f > 10$, we also identified TFs with low Fano factor values, such as Atonal (Ato), Eyeless (Ey) and Spalt Major (Salm).

Ato acts as a differentiation factor for R8 photoreceptor development in the eye disc and its levels are subject to Notch-Delta signaling and auto-regulation (Baker et al., 1996; Jarman et al., 1995). This mechanism might be sufficient to explain the low variation observed.

For Ey, no mechanism of auto-regulation has been determined yet, but cells displaying low or suppressed noise are thought to reside in an undifferentiated state (Kumar et al., 2014), as assessed for pluripotent stem cells. This view is compatible with the low variation observed for Ey in undifferentiated eye disc cells (anterior to the morphogenetic furrow), and the low Fano factor value $F_f \approx 1$ suggests that the

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcription of this gene is likely to be a continuous process that occurs with a constant probability and according to the statistics of a Poisson process.

Interestingly, Salm was the only TF examined for which $F_f < 1$. Given that the distribution of Salm is constrained by the Decapentaplegic (Dpp) signaling pathway that controls growth and patterning during imaginal disc development (Organista et al., 2015), it is perhaps not surprising that Salm displays a pattern of occurrence that is more regular than the randomness associated with a Poisson process.

Finally, we have measured $F_f = 10$ for Sex combs reduced (Scr). This value is equal to the value measured for *Scr* mRNA transcripts in the embryo (Boettiger and Levine, 2013; Pare et al., 2009). Quantification of *Scr* mRNA (and also the ones of *Ubx* and *Deformed* (*Dfd*)) in stage 11 embryos exhibited a very high degree of variation (expressed as standard deviation over the mean) among very closely related cells in the *Scr* expression domain. Although mRNA levels of these TFs may not necessarily reflect protein concentration, comparison of variation in mRNA levels and the corresponding variation in protein concentration of a collection of TFs identified a linear relationship between the two, in a system of simple repression architecture (Sanchez et al., 2011). Taken together, these findings demonstrate that some Hox TFs display increased variation, whether at the level of mRNA or protein concentration, but no generic rule applies (Antp displays comparatively low variation for its concentration).

Our FCS measurements of Antp in diverse tissues, variable concentrations, different developmental time points using distinct isoforms have allowed us to calculate the endogenous apparent dissociation constant for specific Antp-DNA binding to be $K_d^{3rd\ instar} = 18\ nM$ and the average number of Antp molecules per nucleus to be roughly 7000. The measured K_d value is similar to the apparent

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

646 dissociation constant calculated for Scr in salivary gland cells (Vukojevic et al.,
647 2010), $K_{d,s}^{Scr} = 7\,nM$. The similarity of the values could be explained by the fact that
648 these two TFs are, in terms of their homeodomains and hence DNA-binding
649 properties, the most closely related Hox TFs in flies. Conversely, the roughly
650 threefold stronger binding of Scr could be explained by the fact that Scr
651 homodimerizes upon DNA binding, whereas Antp does not (Papadopoulos et al.,
652 2012). Previous studies examined binding of a recombinant Antp homeodomain
653 peptide to an 18 base pair DNA probe of the BS2 binding site
654 (GAGAAAAAGCCATTAGAG) of a *ftz* enhancer. These studies derived dissociation
655 constants in the range of $1 - 2\,nM$ ($K_d = 1.6\,nM$) (Affolter et al., 1990; Muller et al.,
656 1988). The tenfold difference between the in vitro binding constant and the apparent
657 in vivo dissociation constant in the two datasets can be explained by: a) the lack of a
658 full-length protein in the in vitro binding assay, b) interactions with other proteins in
659 the densely packed nuclear environment (including, but not limited to, cofactors),
660 which are not present in the in vitro binding constant, or c) the existence of a
661 collection of binding sites with different affinities for Antp.

662 We have demonstrated that the relative efficiencies of auto-activation and
663 repression rely considerably on the type of Antp isoform engaged in the regulation of
664 the P1 promoter (without excluding additional mechanisms). The size of the linker
665 between the YPWM motif and the homeodomain is important, with short linker
666 isoforms behaving as more efficient repressors than their long linker counterparts.
667 We have previously studied the regulatory capacity (transcriptional activation and
668 repression) of long and short linker Antp isoforms and found that the long linker ones
669 behave as more efficient activators, whereas the short linker ones as effective
670 repressors (Papadopoulos et al., 2011). The observed efficiencies for auto-activation

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and repression are in line with our previous findings. Moreover, we have demonstrated that auto-repression relies on a small portion of the protein, which contains the homeodomain and its C terminus. The YPWM motif, which binds cofactors, such as Extradenticle (Exd), is dispensable for auto-repression. Whether Exd retains some other role in the auto-repression mechanism described remains to be investigated. In contrast, the YPWM motif is required for at least part of the auto-activating function of Antp, since mutations of it did not allow auto-activation to take place, as compared to the wild type TF. Nevertheless, the question remains as to how the linker size (the three-dimensional spacing between the homeodomain and the YPWM motif) could establish preferences on transcriptional activation or repression, even in the absence of a functional YPWM motif (at least for the repressive function). Our study indicates additional functions of the linker region (or parts of it), independent of the YPWM motif. These could be due to interactions with other cofactors or to differential binding affinities of the homeodomain on a subset of binding sites in the presence of a short or a long amino acid stretch.

Throughout this study, we have used the P1 promoter reporter to dissect the regulatory behavior of Antp on its own expression. The reason for this has been that P1 expression patterns have been thoroughly characterized previously in imaginal discs, (Engstrom et al., 1992; Zink et al., 1991) and unpublished results. Whether additional auto-regulation of Antp occurs also from the proximal P2 promoter in discs, remains to be examined, but our initial experiments provided negative evidence of such regulation, taking place in imaginal discs (data not shown). Auto-activation of the *Antp* P2 promoter has previously been suggested to take place in the embryonic nervous system and this mechanism has been invoked to explain why no *Antp* P2 reporter transcription was observed in neuronal cells in *Antp* mutant

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

embryos (Appel and Sakonju, 1993), whereas global auto-regulation of *Hox* gene complexes has been shown to be in effect also in mammalian limb development (Sheth et al., 2014). These pieces of evidence suggest evolutionarily conserved mechanisms for establishing (auto-activation) or limiting (auto-repression) Hox levels and variation in different developmental contexts. Nevertheless, at least in the case of the 17.2 kb P1 reporter, for which auto-regulation was shown in imaginal discs (this study), rigorous *cis*-element analyses should be able to identify relevant enhancers and silencers responsible for Antp auto-regulation. These might well be the regulatory elements identified to function in S2 cultured cells (Saffman and Krasnow, 1994; Winslow et al., 1989), although this hypothesis still needs to be examined.

Negative auto-regulation has been identified as a frequently deployed mechanism for the reduction of noise (cell-to-cell variability) and the increase of regulatory robustness in various systems (Becskei and Serrano, 2000; Dublanche et al., 2006; Gronlund et al., 2013; Nevozhay et al., 2009; Shimoga et al., 2013; Thattai and van Oudenaarden, 2001). Auto-repression has been described for *Ultrabithorax* (*Ubx*) in haltere specification (Crickmore et al., 2009; Garaulet et al., 2008) or in *Ubx* promoter regulation in Drosophila S2 cells (Krasnow et al., 1989), whereas an auto-activating mechanism is responsible for the maintenance of *Deformed* expression in the embryo (Kuziora and McGinnis, 1988). We have established that an auto-inhibitory mechanism, that is analogous to that of *Ubx*, exists for *Antp*. Similar to our observations with Antp, only a small portion of *Ubx*, containing the homeodomain and C-terminus, is sufficient to mediate Ubx repression (Crickmore et al., 2009).

Figure legends

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1: **Measurement of average concentrations of 14 endogenously-tagged TFs in Drosophila imaginal discs by Fluorescence Correlation Spectroscopy (FCS).** (A-P) Fluorescence imaging of TFs, showing their expression pattern in imaginal discs and the salivary gland. White arrows indicate regions where FCS measurements of endogenous intra-nuclear concentration were performed and the average concentrations are given for each TF. Images have been contrasted for visualization purposes. For the Antp and Grn TFs, both leg and wing imaginal discs have been used for measurements. Average concentrations of TFs measured in different cells span a range of two orders of magnitude, from few tens a thousand nanomolar. (Q) Representative temporal autocorrelation curves of eight TFs, selected to demonstrate the full span of the concentration range observed. Note that the amplitude of the autocorrelation curves is inversely proportional to the number of molecules (concentration). (R) Temporal autocorrelation curves shown in (Q) normalized to the same amplitude, $G_n(\tau) = 1$ at $\tau = 10\ \mu s$. Note that the shift of the characteristic decay time to higher values, and the increase of the relative amplitude of the slow component, indicate higher degree of TF binding to the DNA. Scale bars denote $100\ \mu m$, unless otherwise indicated.

Figure 2: **Variation in TF concentration among neighboring nuclei decreases with increasing concentrations.** (A) Characterization of nucleus-to-nucleus variability among neighboring cells within the same expression domain in imaginal discs of the 14 TF studied by FCS. Black bars show concentration averages (with error bars representing 1 standard deviation), whereas grey bars show the variation, i.e. the squared coefficient of variation (expressed as the variance over the squared

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mean, $CV^2 = \frac{s^2}{m^2}$). The Antp measurements are color-coded red for easiness of comparison with (B). (A') Fano factor values ($F_f = \frac{s^2}{m}$) of the 14 TFs for which FCS measurements were performed range from 0-20. (B) Variation of the 14 TFs as a function of concentration. In general, variation is elevated ($CV^2 > 0.2$) and decreases with increasing average TF concentrations. Red squares correspond to measurements of average Antp variation. As in (A), Antp in the wing, but even more in the leg, disc displays markedly lower variation for its average concentrations. Error bars represent 1 standard deviation. (C) Variation in concentration of endogenous Antp in the leg and wing discs, observed among clusters of neighboring cells, each displaying different average concentration, is plotted as a function of concentration. Error bars represent 1 standard deviation. (C') Fano factor values of endogenous Antp as in (B), plotted as a function of concentration. The Fano factor increases with increasing Antp concentrations. Error bars indicate 1 standard deviation.

Figure 3: **Antp is able to repress and activate itself at the transcriptional level.** (A-C') Clonal overexpression of a *SynthAntp-eGFP* construct in the wing disc represses the endogenous Antp protein. Arrows in (B) indicate a clone induced in the notum within the *Antp* expression domain and the dashed line shows the area of repressed Antp. (D-F) Auto-repression of Antp occurs at the transcriptional level. Induction of *SynthAntp-eGFP*-overexpressing clones, while monitoring *Antp* transcription by the *Antp* P1-*lacZ* reporter. Arrows indicate clones within the *Antp* P1 reporter expression domain that express *SynthAntp-eGFP* and down-regulate *Antp* P1-*lacZ*. (G-I') Antp auto-repression is not a dominant loss-of-function phenotype. Clones expressing un-tagged, full-length *Antp*, labeled by cytoplasmic eGFP (white arrows) repress transcription of the reporter. (J-L) Expression of *SynthAntp-eGFP* by

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dll-Gal4 (MD23) results in ectopic activation of *Antp* P1-*lacZ* in the wing disc. Note the endogenous and ectopic expression of the reporter in (B), indicated by white arrows. (M) X-gal staining of *Antp* P1 reporter upon *SynthAntp-eGFP* induction confirms the endogenous and ectopic activity (black arrows) of β-galactosidase in the distal part of the wing disc. (N) Control X-gal staining of *Antp* P1-*lacZ*. Black arrows indicate the endogenous expression. (O-Q) Ectopic activation of the reporter in the distal antennal portion of the eye-antennal disc (white arrow), as in (K). (R) X-gal staining of antennal discs (black arrows) confirms the results of stainings in (F-H). (S) Control antennal discs bearing the *Antp* P1 reporter show no transcription in the eye-antennal disc. (T-V) Stainings of *Antp* P1-*lacZ* in the distal portion of the wing disc upon expression of cytoplasmic eGFP from the *Dll* enhancer show no activity of the P1 promoter (negative control). Scale bars denote $100\ \mu m$.

Figure 3 – figure supplement 1: **Schematic representation of the Antp-eGFP fusion protein produced by the conversion of the MiMIC MI02272 construct to an artificial exon.** The eGFP-encoding artificial exon is situated in intron 6 of the mRNA and is spliced in between exons 6 and 7 that correspond to the long and non-conserved N-terminal coding sequence of the protein, which has little (if any) function *in vivo* (Papadopoulos et al., 2011), and does not disrupt the homeodomain or YPWM motif. All features have been drawn to scale.

Figure 3 – figure supplement 2: **Antp expression patterns are not altered by the MiMIC MI02272 insertion.** Heterozygous flies (embryos and third instar larvae), examined for their Antp-eGFP pattern (detected by an antibody to GFP, green), as compared to the total amount of Antp (expressed by the sum of the MiMIC *Antp*-

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

eGFP and the wild type *Antp* loci), detected by an Antp antibody (magenta). Comparisons of the Antp pattern in wild type embryos and all thoracic imaginal discs are provided case-wise in the right panel. In discs, dashed lines approximately separate the anterior (A) from the posterior (P) domain of the disc. Note the very high expression of Antp in the humeral disc. In the leg discs, Antp is expressed most strongly in the posterior compartment of the prothoracic leg disc, the anterior compartment of the mesothoracic leg disc and in an abundant pattern in the metathoracic leg disc. Cyan arrows point to Antp positive cells in the second and third leg discs that are centrally located, as previously shown (Engstrom et al., 1992). All images represent Z-projections. Scale bars denote $100\ \mu m$.

Figure 3 – figure supplement 3: **Auto-repression of Antp-eGFP in thoracic imaginal discs, assessed by different Gal4-drivers.** (Upper panel) Repression of Antp-eGFP in all ventral thoracic discs (leg discs), using *Dll*- and *ptc*-Gal4. eGFP has been used as a negative control throughout. (Lower panel) Repression of the *Antp* P1-*lacZ* reporter in the leg disc (exemplified by a prothoracic leg disc), upon overexpression of a SynthAntp from the MD23 *Dll* enhancer. White arrows in the upper and lower panels point at regions of overlap between the Gal4-driver and the *Antp* expression domains, where repression is observed (SynthAntp) or is not detected (eGFP). Scale bars denote $100\ \mu m$.

Figure 3 – figure supplement 4: **Antp is able to auto-activate its transcription independently of Dll.** (A-C') Antp repression and ectopic activation in the wing disc is independent of Dll. Clonal overexpression of SynthAntp-eGFP in the prescutum region of the notum, upon knockdown of *Dll* by *RNAi* shows persistence of

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

821 repression of Antp (white arrows), whereas clones in the disc outside of the *Antp*
822 expression domain, but within the *Dll* expression domain show no ectopic activation
823 of Antp (yellow arrows), precluding that the ectopic activation in Figure 3 (J-M and O-
824 R) is caused by a *Dll* sensitized background. However, no detectable ectopic
825 activation is triggered elsewhere in the disc (cyan arrows), as should have been
826 expected (see text for explanation). (D-E) Clonal overexpression of cytoplasmic
827 eGFP upon *Dll* knockdown does not interfere with *Antp* expression. Scale bars
828 denote $100\ \mu m$.
829
830 Figure 4: **Antp is sufficient and required to trigger a developmental switch from**
831 **transcriptional auto-activation to auto-repression.** (A) Schematic representation
832 of the experimental setup. Embryos were collected and allowed to develop for 26 h
833 (early clones) or 60 h (late clones), at which time overexpression clones were
834 induced. They were then harvested at third instar wandering larval stage for analysis.
835 (B-D) Early clones, expressing full-length, untagged *Antp*, marked by the absence of
836 mCherry, reveal strong auto-activation of *Antp-eGFP* expression (red arrows in (B)
837 and (D)), monitored by the endogenous Antp-eGFP protein. (E-G) Close-up of the
838 notum region of the prescutum where endogenous *Antp-eGFP* is strongly expressed.
839 Dashed lines in (E) and (G) delineate the region of strong endogenous expression.
840 Note the absence of repression of the endogenous gene in overexpression clones
841 within this region at this stage. The whole *Antp* expression domain expresses *Antp-*
842 *eGFP*, but overexpression clones (sub-regions of absence of mCherry staining)
843 express *Antp-eGFP* even stronger. (H-J) Late clones overexpressing untagged, full-
844 length Antp repress endogenous *Antp*, (white arrows in (H) and (J)). (K-R') Antp is
845 required for sustaining its own expression (auto-activation) and for limiting its

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression levels (auto-repression) during imaginal disc development. (K-L') In positive control discs, clonal knockdown of Antp in the wing disc results in efficient down-regulation of the *Antp-eGFP* allele. (M-N') Induction of *RNAi* knockdown clones of *Antp* at 26 h of development (early clonal induction) results in cells lacking *Antp* P1 reporter expression. (O-P) Induction of *Antp* knockdown clones by *RNAi* at later stages (i.e. 60 h of development) within its normal expression domain results in up-regulation of the *Antp* P1 reporter above endogenous levels. In panels K-P, nuclear mRFP1 marks the cells of *Antp* knockdown by *RNAi*. (Q-R') Representative control disc, in which clonal overexpression of *eGFP* (following either of the two clone-induction regimes, at 26 h or 60 h of development) has been induced, and which results in no reduction or increase of Antp protein. Scale bars denote $100\ \mu m$, unless otherwise indicated.

Figure 4 – figure supplement 1: **Early auto-activation of *Antp* occurs at the transcriptional level.** Early induction (26 h of development) of SynthAntp in the wing, leg and eye-antennal disc results in strong ectopic activation of the *Antp* P1 reporter. Dashed white lines outline the domains of endogenous expression in the wing disc, whereas yellow arrows point at clones, which show ectopically activated *Antp* P1 transcription. Scale bars denote $100\ \mu m$.

Figure 5: **Direct correlation between Antp concentration and homeotic function – Antp auto-repression and activation occurs at endogenous concentrations.** (A-D) Live imaging (one optical section) of SynthAntp-eGFP expressed in the distal antennal portion of the eye-antennal disc, induced by different Gal4 drivers. The concentration was measured using FCS and average concentrations are indicated.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

An eightfold difference was observed between the strong *Dll*-Gal4 driver (MD23) and weak *69B*-Gal4 driver. (E-H) Transformations of the distal antenna into a tarsus in adult flies, caused by *SynthAntp-eGFP* overexpression in antennal discs (A-D) The strength of the transformation correlates with the level of expression from the different Gal4 drivers. Ectopic tarsi range from complete transformation to milder transformations of the arista or ectopic leg bristles in the third antennal segment in (G) and (H), indicated by black arrows. (I-I') Representative temporal autocorrelation curves in nuclei overexpressing *SynthAntp-eGFP* (I) and comparison between average concentrations, represented by the number of molecules in the OVE, obtained using different Gal4 drivers (I'). Note that the amplitude of the autocorrelation curve is inversely proportional to the concentration. Statistical significance was determined using a two-tailed Student's T-test (***, p<0.001, namely $p_{MD23-ptc} = 3.3 \cdot 10^{-3}$, $p_{MD23-MD713} = 7.2 \cdot 10^{-4}$, $p_{MD23-69B} = 6 \cdot 10^{-6}$, $p_{ptc-69B} = 5.8 \cdot 10^{-14}$ and $p_{MD713-69B} = 3 \cdot 10^{-15}$; n.s., non-significant, namely $p_{ptc-MD713} = 5.9 \cdot 10^{-2}$). (J) Temporal autocorrelation curves of measurements in (I) normalized to the same amplitude, $G_n(\tau) = 1$ at $\tau = 10\,\mu s$, show major overlap, indicating indistinguishable behavior of Antp binding to the DNA across the concentration range examined $(0.5 - 3.8\,nM)$. (K-L) Repression of endogenous Antp protein upon induction of *SynthAntp-eGFP* in the proximal regions of the wing disc by *69B*-Gal4. (M-N) No repression is observed upon overexpression of *eGFP* (negative control). (O) X-gal stainings of the *Antp* P1 reporter show weak but definite ectopic β-galactosidase activity in the antennal disc (black arrows). (P) Control stainings of eGFP induced by the *69B* enhancer show complete absence of ectopic reporter transcription. Scale bars denote $100\,\mu m$, unless otherwise indicated.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5 – figure supplement 1: **Ectopic Antp auto-activation correlates with Antp concentration.** (A-D) Ectopic activation of the *Antp* P1 transcriptional reporter upon expression of *SynthAntp-eGFP* in the pouch region of the wing disc (A-B) and the antennal portion of the eye-antennal disc (C-D) by the MD713 *Dll*-Gal4 driver. Weak, but definite ectopic activation is observed. (E-F') Widespread overexpression of SynthAntp from the *69B* enhancer results in repression of endogenous Antp in the proximal thoracic structures of the wing disc (white arrows), but no ectopic activation is detectable with this assay (see text for explanation). (G-H') Auto-repression of endogenous Antp in the wing disc is detected upon ectopic expression of *SynthAntp* by *ptc*-Gal4 in the region of overlap between the *ptc* and the *Antp* expression domains (white arrow in (G')). Additionally, weak auto-activation of Antp is detected along the anterior-posterior margin (cyan arrows in (G')). (I-J') Neither repression nor activation occur when eGFP alone is induced by *ptc*-Gal4 (negative control). Scale bars denote $100\ \mu m$.

Figure 6: **Antp transitions to a state of higher concentration, lower variation and longer DNA residence times during larval disc development – study of endogenous Antp-DNA binding by FCS.** (A-B) Live imaging of *Antp-eGFP* expressing nuclei in second instar leg (A) and wing (B) imaginal discs. (C) Quantification of average concentrations and cell-to-cell variability in protein concentration among neighboring nuclei in wing and leg, second and third instar, discs. Black bars denote the average concentration and red bars denote the squared coefficient of variation in concentration, expressed as the variance over the squared mean. Note the increase in average concentration from second to third instar (eleven-fold increase in the leg disc) and the concurrent drop in variation to almost

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

half of its value. Statistical significance was determined using Student's two-tailed T-test (***, p<0.001, namely $p_{3rd-2nd\ instar\ leg} = 4.4 \cdot 10^{-18}$ and $p_{3rd-2nd\ instar\ wing} = 3.2 \cdot 10^{-8}$). (D) Representative temporal autocorrelation curves recorded in second and third instar wing and leg imaginal discs, expressing *Antp-eGFP*. Note the low concentration in second instar leg and wing discs, reflected by the relatively high amplitude of the autocorrelation curves (as it is inversely proportional to the number of molecules), as compared to the high concentration in third instar discs. (E) Temporal autocorrelation curves shown in (D) normalized to the same amplitude, $G_n(\tau) = 1$ at $\tau = 10\ \mu s$, show both a shift towards longer characteristic decay times in the third instar leg and wing discs and an increase in the relative amplitude of the second component, indicative of the increased fraction of TF implicated in interactions with the DNA. Autocorrelation curves are color-coded as outlined in panel (D). (F) Autocorrelation curves of Antp-eGFP in wing disc nuclei. Concentration differences of fluorescent Antp protein are obvious between cells expressing one or two copies of Antp-eGFP (homozygous and heterozygous larvae). A two-component model for free three-dimensional diffusion and triplet formation has been used for fitting (black lines). (G) Autocorrelation curves shown in (F) normalized to the same amplitude, $G_n(\tau) = 1$ at $\tau = 10\ \mu s$, show pronounced overlap between homozygous and heterozygous *Antp-eGFP*-expressing cells, as well as between endogenously expressed *Antp* and overexpressed *SynthAntp-eGFP*. Autocorrelation curves are color-coded as outlined in panel (F). (H) Live imaging of a representative homozygous, *Antp-eGFP*-expressing wing imaginal disc, in which FCS measurements in (F-G) have been performed. *Antp* is expressed in the primordia of the prescutum and the proximal mesothoracic structures. (I) Live imaging of a representative *Antp-eGFP*-expressing homozygous metathoracic leg disc, in which

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FCS measurements in (D-E) have been performed. *Antp* is expressed in the proximal regions of the disc and not restricted to anterior or posterior compartments. (J) Live imaging of the antennal portion of an eye-antennal disc expressing one copy of a synthetic Antp-eGFP construct, as in (Papadopoulos et al., 2011), from the *Dll*-enhancer (MD23). The same optical settings were used as in (H) and (I). (K) Live imaging of the region of maximum *Antp* expression in (H). (L) Close-up of (K). Note the uneven distribution of Antp in the nuclei, its exclusion from the nucleoli (read arrowheads) and the presence of bright spots (sites of pronounced accumulation, yellow arrowheads). All images represent one optical section. (M) Binding study of endogenous Antp-eGFP in third instar wing and leg discs. The concentration of DNA-bound fraction of Antp-eGFP molecules, measured by FCS in (F) and (G), is plotted as a function of the total concentration of Antp-eGFP molecules. The slope and the intercept of the linear regression equation, $y = 0.38x - 5.78$, were used to determine the apparent dissociation constant $K_{d,app.}^{Antp-eGFP} = 18nM$ (for the calculation refer to Supplement 3). (N) Binding study of endogenous Antp-eGFP (third instar wing and leg discs) and overexpressed *SynthAntp-eGFP* (antennal discs by *ptc*-, *Dll* (MD23)-, *Dll* (MD713)-, and *69B*-Gal4) using FCS measurements. From the linear regression equation, $y = 0.46x - 7.29$, the apparent dissociation constant was calculated to be $K_{d,app.}^{Antp-eGFP} = 15nM$ (for the calculation refer to Supplement 5) and is very similar to the value obtained in (M). Note that in both (M) and (N) Antp binding to the DNA increases linearly with increasing concentration of TF, but in the concentration range of this study ($0.2 - 6\ \mu M$), saturation is not reached. (O) Binding study of endogenous Antp-eGFP in second instar wing and leg discs. From the linear regression equation, $y = 0.44x - 2.6$, the apparent dissociation constant was calculated to be $K_{d,app.}^{Antp-eGFP} = 131nM$ (for the calculation refer to Supplement 6).

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

This value is higher that in third instar, indicating weaker affinity of binding to the DNA. Scale bars denote $100\ \mu m$, unless otherwise indicated.

Figure 6 figure supplement 1: **Analysis of characteristic decay times of Antp-eGFP as a function of total concentration and study of Antp binding in third instar discs by Fluorescence Recovery After Photobeaching (FRAP).** (A-B) Characteristic decay times τ_{D_1} (A) and τ_{D_2} (B) do not vary with the concentration of Antp-eGFP TF molecules, as evident from $\tau_{D_1} = f\ (N_1)$ and $\tau_{D_2} = f\ (N_2)$, where N_1 is the number of freely diffusing, N_2 the number of bound Antp-eGFP TF molecules and τ_{D_1}, τ_{D_2} their respective diffusion times. (C) FRAP analysis of endogenous Antp-eGFP in nuclei of wing disc cells. Absence of recovery even after 360 sec (6 min) indicates long residence of endogenous Antp-eGFP on the DNA with $k_{off}^{specific} < 6 \cdot 10^{-3}\ s^{-1}$.

Figure 6 – figure supplement 2: **Binding study of Antp.** (A-C) Close-up of linear regression curves of Antp binding under different conditions to denote the obtained slopes and intercepts that were used in the calculation of the apparent equilibrium dissociation constants.

Figure 7: **Control of relative isoform abundance favors the transition of Antp from auto-activation to auto-repression and is independent of the YPWM motif.** (A) Schematic representation of the linker region between the YPWM motif and the homeodomain. A four amino acid-encoding exon (GKCQ) is spliced 5' to the homeodomain-encoding exon, generating a long-linker isoform. In the mutated SynthAntp constructs the YPWM motif has been substituted by alanines. (B-I') The

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

non-conserved N terminus and the YPWM motif are dispensable for Antp auto-repression, but short linker isoforms display stronger auto-repression capacity. *SynthAntp-eGFP* transgenes encoding Antp isoforms with a wild type YPWM motif and a long (B-C') or a short (D-E') linker, as well as isoforms bearing a substitution of the YPWM motif by alanines and a long (F-G') or a short (H-I') linker were induced from the *ptc* enhancer and Antp protein repression was monitored at the proximal portion of the wing disc (white arrows). (J) The efficiency of the repression was monitored and normalized to the intensity of Antp staining when an eGFP transgene was overexpressed by *ptc*-Gal4. All constructs repress the endogenous Antp protein in various degrees and the auto-repression does not require the YPWM motif. Short linker Antp isoforms repress more efficiently whether in the context of a wild type or mutated YPWM motif. Intensities were calculated using Fiji (Schindelin et al., 2012). Statistical significance was determined using a two-tailed Student's T-test (***, p<0.001, namely $p_{YPWM,long-short\ linker} = 2.4 \cdot 10^{-4}$ and $p_{AAAA,long-short\ linker} = 2.3 \cdot 10^{-7}$; n.s., non-significant, namely $p_{YPWM-AAAA,short\ linker} = 2.7 \cdot 10^{-1}$). (K) Quantification of the relative concentrations of short and long linker *Antp* transcripts in wild type flies at different developmental stages. The relative abundance of the long linker variant, as compared to the short linker one, is higher in the embryo, but gradually decreases during larval stages. In third instar larvae the short linker transcript comprises about 80% of the total *Antp* mRNA. Note that the quantification is relative and not absolute. The significance of differences in repression capacity and relative abundance of long and short linker transcripts in (I-J) have been determined using Student's two-tailed T-test (***, p<0.001, namely $p_{3rd\ instar\ wing-2nd\ instar} = 2.3 \cdot 10^{-5}$, $p_{3rd\ instar\ leg-2nd\ instar} = 8 \cdot 10^{-6}$, and $p_{2nd\ instar-embryo} = 5.5 \cdot 10^{-5}$; n.s., non-significant, namely $p_{3rd\ instar\ leg-wing} = 6.4 \cdot$

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10^{-1}). (L-L') FCS analysis performed on third instar wing and antennal imaginal discs, expressing short or long linker Antp isoforms (tagged to eGFP) close to endogenous concentrations, from the *69B*-enhancer (also studied in Figure 5 D). Cell nuclei of similar concentrations in the two datasets have been selected for analysis (L'). Temporal autocorrelation curves from FCS measurements on the short linker Antp isoform display a consistent shift towards longer characteristic decay times and higher amplitude of the slow component, as compared to its long linker counterpart (L). (M-N) Binding study of short and long linker Antp isoforms in third instar wing and leg discs, expressed by *69B*-Gal4, as in (L-L'). The concentration of the DNA-bound complexes (L-L'), is plotted as a function of the total concentration of Antp-eGFP molecules. From the linear regression equation, $y = 0.34x - 5.31$, the apparent dissociation constant for the short linker isoform was calculated to be $K_{d,Antp}^{short\ linker\ isof.,o/e} = 18nM$ (M) and from the linear regression equation $y = 0.24x - 3.28$, the apparent dissociation constant for the long linker isoform was calculated to be $K_{d,Antp}^{long\ linker\ isof.,o/e} = 190nM$ (N) (for the calculation refer to Supplement 6). The two dissociation constants differ by about one order of magnitude, indicating stronger binding of the short linker isoform to the DNA, as compared to the long linker one. Scale bars denote $100\ \mu m$.

<u>Figure 7 – figure supplement 1:</u> **Binding study of Antp.** (A-B) Close-up of linear regression curves of Antp binding under different conditions to denote the obtained slopes and intercepts that were used in the calculation of the apparent equilibrium dissociation constants.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 7 – figure supplement 2: **The Antp YPWM motif is required for Antp auto-activation, but long linker isoforms display stronger auto-activation potential.** (A-D) Overexpression of full-length, long linker *Antp* variant by ptc-Gal4 in the embryo results in weak ectopic activation of the *Antp* P1 reporter in regions anterior and posterior to the *Antp* normal expression domain (A-B), whereas expression of a short linker, full-length variant results in little, if any, ectopic transcription of the reporter (C-D). (E-G) Mutated full-length Antp with conversion of the YPWM motif to alanines completely fails to ectopically activate the reporter (O-P) and shows similar staining as the control embryo in (G). White brackets indicate the normal expression domain of *Antp* in thoracic segments and red dashed lines outline the ectopic expression. Scale bars denote $100\ \mu m$.

Figure 8: **Antp suppresses its variation through a developmental switch from auto-activation to auto-repression that depends on concentration, relative abundance of isoforms and their DNA-binding properties.** Schematic representation of the regulatory role of Antp during the transition from early to late larval thoracic disc development. During early stages, the average concentrations of Antp are relatively low (for simplicity demonstrated here as only a few molecules per nucleus), but the variation in concentration among neighboring cells is relatively high (expressed by the coefficient of variation or the Fano factor). Nuclei containing higher amount of preferentially activating isoforms (featuring a long linker, as shown in Figure 7 A) in the majority of the cells (but not necessarily all of them) undergo transcriptional auto-activation of the *Antp* gene. The binding of the activating isoform to the DNA involves fast diffusion and the affinity to the DNA, evidenced from the apparent equilibrium dissociation constant (K_d), is relatively low. As development

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proceeds, transcriptional auto-activation has increased the average concentration of Antp. Subject to an unknown regulatory mechanism (splicing regulation, differential lifetime of mRNA or protein, preferential degradation etc.), the equilibrium of activating and repressing Antp isoforms (featuring a short linker, as shown in Figure 7 A) is shifted, favoring the persistence of the repressing isoforms, which display higher affinity of DNA binding and slower diffusion on the DNA. This regulatory switch accounts for a progressive deceleration/cessation of the production of more Antp protein and suppresses cell-to-cell variability.

Figure 8 – figure supplement 1: **Maintenance of Antp expression is required for normal leg and notum development.** (A) Gain- and loss-of-function phenotypes generated by overexpression of *SynthAntp-eGFP* (upper row) and *mCitrine-SynthScr* (lower row) by the *Dll*-Gal4 (MD23) driver. While flies expressing *SynthAntp* are fully viable and all display ectopic tarsi in the antenna, flies expressing SynthScr also display tarsal transformations in the antenna, but have complete absence of tarsal segments in the legs and die as pharate adults. Black arrowheads indicate the antennal tarsi and red arrows the malformed legs. (B-C) Repression of endogenous Antp by induction of *SynthScr* from the *ptc* enhancer. (D-E) Repression of endogenous *Antp* by SynthScr takes place at the transcriptional level, as visualized by the *Antp* P1 reporter. The arrows in (B) and (D) indicate the region of Antp repression, where the *ptc* and *Antp* expression domains overlap. (F-G) Contrary to SynthAntp, SynthScr fails to ectopically activate the *Antp* P1 reporter in the wing disc, despite overexpression by the “strong” *Dll*-Gal4 driver (MD23). (H-I) Contrary to Antp, SynthScr fails to repress itself in the prothoracic leg disc, when overexpressed by *Dll*-Gal4 (MD23). Dashed line marks the region of major overlap between

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

endogenous and SynthScr proteins. (J-K') Repression of endogenous Antp in the notum, caused by ectopic expression of *SynthScr* from the *MS243*-Gal4 driver. White brackets in (J) and (J') show the region of Antp repression in the notum. (L-M) Ectopic expression of *SynthAntp* by *MS243*-Gal4 causes mild defects in the scutelum (indicated by black arrows), whereas the remaining notum structures are wild type-like (L). The observed eye-reduction is caused by a *Dr* allele. In contrast, ectopic expression of *SynthScr* causes severe malformations of the notum and head, including complete loss of eyes (M). (N-Q) Control discs, showing that ectopic expression of *SynthAntp* by *MS243*-Gal4 causes repression of endogenous Antp in the notum (white arrows) (N-O'), but no repression is observed by ectopic expression of *eGFP* alone (P-Q). Scale bars denote $100\ \mu m$, unless otherwise indicated.

Acknowledgements

We are deeply saddened by the unexpected passing of Prof. Walter J. Gehring, at the very inception of this work, when the project was still in the planning and preliminary data gathering stage. Prof. Gehring was an extraordinary human being and a scientific giant, whose work will continue to educate and inspire generations to come. The authors are indepted to Sonal Nagarkar Jaiswal, Paolo Mangahas, and Hugo J. Bellen for creating and sharing with us the Antp-eGFP line. DKP has been supported by a long-term fellowship from the Swiss National Science Foundation (PBBSP-138700) and a long-term fellowship from the Federation of European Biochemical Societies (FEBS) at initial stages of this project. VV has been supported by the Knut and Alice Wallenberg foundation and Karolinska Institute Research Funds. DKP would like to express his gratitude to PT for outstanding

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

scientific, and uninterrupted financial, support. DKP is indebted to Markus Burkhardt, head of the imaging platform at the Center for Regenerative Therapies Dresden (CRTD), for help and discussions regarding FCS experiments in Dresden; Sylke Winkler and the DNA sequencing facility of the Max-Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG) for providing assistance with Antp transcript variant quantification; as well as the Light Microscopy facility of MPI-CBG. DKP is also grateful to KS for numerous discussions and support throughout the implementation of this work.

Materials and Methods

Fly stocks used

The Antp-eGFP MiMIC line has been a kind gift from Hugo J. Bellen. The *atonal* (VDRC ID 318959), *brinker* (VDRC ID 318246), *spalt major* (VDRC ID 318068), *yorkie* (VDRC ID 318237), *senseless* (VDRC ID 318017) and *Sex combs reduced* (VDRC ID 318441) fosmid lines are available from the Vienna Drosophila Resource Center (VDRC) and have been generated recently in our laboratory (Sarov et al., 2016). The *fork head* (stock 43951), *grainy head* (stock 42272), *Abdominal B* (stock 38625), *eyeless,* (stock 42271), *spineless* (transcript variant A, stock 42289), and *grain* (stock 58483) tagged BACs were generated by Rebecca Spokony and Kevin P. White and are available at the Bloomington Stock Center. For the *scalloped* gene, a GFP-trap line was used (Buszczak et al., 2007), a kind gift from Allan C. Spradling laboratory (line CA07575), with which genome-wide chromatin immunoprecipitation experiments have been performed (Slattery et al., 2013). For the *spineless* gene, Bloomington stock 42676, which tags isoforms C and D of the

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Spineless protein has been also tried in fluorescence imaging and FCS experiments, but did not yield detectable fluorescence in the antennal disc, rendering it inappropriate to be used in our analysis. Therefore, we resided to stock 42289, which tags the A isoform of the protein. For the *eyeless* gene, the FlyFos015860(pRedFlp-Hgr)(ey13630::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT line (VDRC ID 318018) has been tried also in fluorescence imaging and FCS experiments, but did not yield detectable fluorescence in the eye disc for it to be used in our analysis. The *act5C*-FRT-yellow-FRT-Gal4 line used for clonal overexpression or knock down has been described (Ito et al., 1997). The UAS-Antp lines (synthetic and full-length, wild type or modified), as well as UAS-Scr constructs have been previously described (Papadopoulos et al., 2011; Papadopoulos et al., 2010). The *Dll*-Gal4 (MD23 and MD713) lines have been a kind gift of Ginés Morata (Calleja et al., 1996). *69B*-Gal4 and *ptc*-Gal4 have been obtained from the Bloomington Stock Center. The *Antp* P1-*lacZ* has been previously described (Engstrom et al., 1992; Zink et al., 1991) and spans the region between 9.4 kb upstream of the P1 promoter transcription initiation site and 7.8 kb downstream into the first intron, including the first exon sequences and thus comprising 17.2 kb of Antp regulatory sequences (pAPT 1.8). The line used has been an insertion of the pAPT 1.8 vector bearing the P1 promoter regulatory sequences upstream of an *actin-lacZ* cytoplasmic reporter and has been inserted in cytogenetic location 99F on the right chromosomal arm of chromosome 3. The *Dll*- and *Antp*-RNAi lines have been from VDRC, lines GD4607 and KK101774, respectively. UAS-eGFP stock was a kind gift of Konrad Basler. We are indebted to Sebastian Dunst for generating the *ubi*-FRT-mCherry(stop)-FRT-Gal4(VK37)/CyO line, which drives clonal overexpression upon flippase excision, while simultaneously marking cells by the

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

loss of mCherry. For red-color labeling of clones the *act5C*-FRT-CD2-FRT-Gal4, UAS-mRFP1(NLS)/TM3 stock 30588 from the Bloomington Stock Center has been used. The *MS243*-Gal4; UAS-GFP/CyO line was a kind gift from the laboratory of Ernesto Sánchez-Herrero.

Fly genotypes of fluorescence images

Figure 1 A: FlyFos018487(pRedFlp-Hgr)(ato37785::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT

Figure 1 B: FlyFos024884(pRedFlp-Hgr)(brk25146::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT

Figure 1 C: FlyFos030836(pRedFlp-Hgr)(salm30926::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT

Figure 1 D: FlyFos029681(pRedFlp-Hgr)(yki19975::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT

Figure 1 E: w^{1118}; PBac(fkh-GFP.FPTB)VK00037/SM5

Figure 1 F: *sd*-eGFP (FlyTrap, homozygous)

Figure 1 G: w^{1118}; PBac(grh-GFP.FPTB)VK00033

Figure 1 H: FlyFos018974(pRedFlp-Hgr)(Scr19370::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT

Figure 1 I: FlyFos015942(pRedFlp-Hgr)(sens31022::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT

Figure 1 J and K: Antp-eGFP (MiMIC) homozygous (line MI02272, converted to an artificial exon)

Figure 1 L: w^{1118}; PBac(Abd-B-EGFP.S)VK00037/SM5

Figure 1 M: w^{1118}; PBac(ey-GFP.FPTB)VK00033

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1 N: w^{1118}; PBac(ss-GFP.A.FPTB)VK00037

Fi.g 1 O and P: w^{1118}; PBac(grn-GFP.FPTB)VK00037

Figure 3 A-C': *hs*-flp; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-SynthAntp-eGFP/+

Figure 3 D-F: *hs*-flp; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-SynthAntp-eGFP/*Antp* P1-*lacZ*

Figure 3 G-I': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4, UAS-eGFP/+; UAS-Antp (full-length, untagged)/*Antp* P1-*lacZ*

Figure 3 J-M and O-R: *Dll*-Gal4 (MD23)/+; UAS-SynthAntp-eGFP/*Antp* P1-*lacZ*

Fig 3 N, S and T-V: *Dll*-Gal4 (MD23)/+; UAS-eGFP/*Antp* P1-*lacZ*

Figure 4 B-J: *hs*-flp/+; *ubi*-FRT-mChery-FRT-Gal4/+; Antp-eGFP (MiMIC)/UAS-Antp (full-length, untagged)

Figure 4 K-L': *hs*-flp/+; UAS-Antp-RNAi/+; Antp-eGFP (MiMIC)/*act5C*-FRT-CD2-FRT-Gal4, UASmRFP1(NLS)

Figure 4 M-P: *hs*-flp/+; UAS-Antp-RNAi/+; *Antp* P1-*lacZ*/*act5C*-FRT-CD2-FRT-Gal4, UASmRFP1(NLS)

Figure 4 Q-R': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4/+; *Antp* P1-*lacZ*/UAS-eGFP

Figure 5 A and E: *Dll*-Gal4 (MD23)/+; UAS-SynthAntp-eGFP/+

Fig 5 B and F: *ptc*-Gal4/+; UAS-SynthAntp-eGFP/+

Figure 5 C and G: *Dll*-Gal4 (MD713)/+; UAS-SynthAntp-eGFP/+

Figure 5 D, H and K, L, O: *69B*-Gal4/UAS-SynthAntp-eGFP

Figure 5 M, N and P: *69B*-Gal4/UAS- eGFP

Figure 6 A, B, H, I, K and L: Antp-eGFP (MiMIC) homozygous

Figure 6 J: *Dll*-Gal4 (MD23)/+; UAS-SynthAntp-eGFP/+

Figure 7 B-C': *ptc*-Gal4/+; UAS-Synth^{YPWM,LL}Antp-eGFP/+

Figure 7 D-E': *ptc*-Gal4/+; UAS-Synth^{YPWM,SL}Antp-eGFP/+

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 7 F-G': *ptc*-Gal4/+; UAS-Synth^{AAAA,LL}Antp-eGFP/+

Figure 7 H-I': *ptc*-Gal4/+; UAS-Synth^{AAAA,SL}Antp-eGFP/+

Figure S2: Antp-eGFP (MiMIC)/TM3, *dfd*-YFP

Figure S3 upper panel: *Dll*-Gal4 (MD23)/+; UAS-SynthAntp-eGFP/+ or *ptc*-Gal4/+; UAS-SynthAntp-eGFP/+ and *Dll*-Gal4 (MD23)/+; UAS- eGFP/+ or *ptc*-Gal4/+; UAS-eGFP/+ in control stainings

Figure S3 lower panel: *Dll*-Gal4 (MD23)/+; UAS-SynthAntp-eGFP/*Antp* P1-*lacZ* and *Dll*-Gal4 (MD23)/+; UAS- eGFP/*Antp* P1-*lacZ* in control stainings

Figure S4 A-C'': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-SynthAntp-eGFP/UAS-*Dll*-RNAi

Figure S4 D and E: *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-eGFP/UAS-*Dll*-RNAi

Figure S5: *hs*-flp; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-SynthAntp-eGFP/*Antp* P1-*lacZ*

Figure S6 A-D: *Dll*-Gal4 (MD713)/+; UAS-SynthAntp-eGFP/*Antp* P1-*lacZ*

Figure S6 E-F': *69B*-Gal4/UAS-SynthAntp-eGFP

Figure S6 G-H': *ptc*-Gal4/+; UAS-SynthAntp-eGFP/+

Figure S6 I-J': *ptc*-Gal4/+; UAS-eGFP/+

Fig S9 A and B: *ptc*-Gal4/+; *Antp* P1-*lacZ*/UAS-^{YPWM,LL}Antp (full-length, untagged)

Figure S9 C and D: *ptc*-Gal4/+; *Antp* P1-*lacZ*/UAS-^{YPWM,SL}Antp (full-length, untagged)

Figure S9 E and F: *ptc*-Gal4/+; *Antp* P1-*lacZ*/UAS-^{AAAA,LL}Antp (full-length, untagged)

Figure S9 G: *Antp* P1-*lacZ*/TM3, *dfd*-YFP

Figure S10 A upper row: *Dll*-Gal4 (MD23)/+; UAS-SynthAntp-eGFP/+

Figure S10 A lower row, B-C', H and I: *Dll*-Gal4 (MD23)/+; UAS-mCitrine-SynthScr/+

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S10 D-E: *ptc*-Gal4/+; UAS-mCitrine-SynthScr/*Antp* P1-*lacZ*

Figure S10 F and G: *Dll*-Gal4 (MD23)/+; UAS-mCitrine-SynthScr/*Antp* P1-*lacZ*

Figure S10 J-K' and M: *MS243*-Gal4/+; UAS-mCitrine-SynthScr/Dr

Figure S10 L and N-O': *MS243*-Gal4/+; UAS-SynthAntp-eGFP/Dr

Figure S10 P and Q: *MS243*-Gal4; UAS-eGFP

Preparation of second and third instar imaginal discs for FCS measurements.

For FCS measurements, imaginal discs (eye-antennal, wing, leg, humeral and genital) and salivary glands were dissected from third instar wandering larvae, or wing and leg discs from second instar larvae, in Grace's insect tissue culture medium (ThermoFisher Scientific, 11595030) and transferred to 8-well chambered coverglass (Nunc® Lab-Tek™, 155411) containing PBS just prior to imaging or FCS measurements. Floating imaginal discs or salivary glands were sunk to the bottom of the well using forceps.

Immunostainings in embryos and larval imaginal discs

Embryonic stainings were performed using the precise protocol from (Grieder et al., 2000). Larval imaginal discs were stained according to (Papadopoulos et al., 2010). Stainings for the endogenous Antp protein have been performed using a mouse anti-Antp antibody (Developmental Studies Hybridoma Bank, University of Iowa, anti-Antp 4C3) in a dilution of 1:250 for embryos and 1:500 for imaginal discs. eGFP, or eGFP-tagged proteins have been stained using mouse or rabbit anti-GFP antibodies from ThermoFisher Scientific in a dilution of 1:500 in imaginal discs and 1:250 in embryos. mRFP1 was stained using a Chromotek rat anti-RFP antibody. A rabbit anti-β-galactosidase antibody (Cappel) has been used for *Antp* P1 promoter

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

stainings in a dilution of 1:1000 in the embryo, whereas for stainings in imaginal discs we used the mouse anti-β-galactosidase 40-1a antibody from Developmental Studies Hybridoma Bank, University of Iowa in a dilution of 1:50. Confocal images of antibody stainings represent predominatly Z-projections and Zeiss LSM510, Zeiss LSM700 or Zeiss LSM880 Airyscan confocal laser scanning microscopy systems with an inverted stand Axio Observer microscope were used for imaging. Image processing has been performed in Fiji (Schindelin et al., 2012). For optimal spectral separation, secondary antibodies coupled to Alexa405, Alexa488, Alexa594 and Cy5 (ThermoFischer Scientific) were used.

Colocalization of wild type and eGFP-tagged MiMIC Antp alleles in imaginal discs

To examine whether the pattern of the MiMIC Antp-eGFP fusion protein recapitulates the Antp wild type expression pattern in both embryo and larval imaginal discs, we performed immunostainings of heterozygous Antp-eGFP and wild type flies to visualize the embryonic (stage 13) and larval expression of *Antp* and eGFP. In this experiment, we 1) visualized the overlap between eGFP and *Antp* (the eGFP pattern reflects the protein encoded by the MiMIC allele, whereas the *Antp* pattern reflects the sum of protein produced by the MiMIC allele and the allele of the balancer chromosome) and 2) compared the eGFP expression pattern to the Antp expression pattern in wild type discs and embryos.

Induction of early and late overexpression and *RNAi*-knockdown clones in imaginal discs

Genetic crosses with approximately 100 virgin female and 100 male flies were set up in bottles and the flies were allowed to mate for 2 days. Then, they were

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transferred to new bottles and embryos were collected for 6 hours at 25°C. Flies were then transferred to fresh bottles and kept until the next collection at 18°C. To asses Antp auto-activation, the collected eggs were allowed to grow at 25°C for 26 h from the midpoint of collection, when they were subjected to heat-shock by submersion to a water-bath of 38°C for 30 min and then placed back at 25°C until they reached the stage of third instar wandering larvae, when they were collected for dissection, fixation and staining with antibodies. To assess Antp auto-repression, the same procedure was followed, except that the heat-shock was performed at 60 h of development after the midpoint of embryo collection. Whenever necessary, larval genotypes were selected under a dissection stereomicroscope with green and red fluorescence filters on the basis of *deformed* (*dfd*)-YFP bearing balancer chromosomes (Le et al., 2006) and visual inspection of fluorescence in imaginal discs.

Measurement of Antp transcript variant abundance

The linker between the Antp YPWM motif and the homeodomain encodes the sequence RSQFGKCQE. Short linker isoforms encode the sequence RSQFE, whereas long linker isoforms are generated by alternative splicing of a 12 base pair sequence encoding the four amino acid sequence GKCQ into the mRNA. We initially designed primer pairs for RT-qPCR experiments to distinguish between the short and long linker mRNA variants. For the short linker variant, we used nucleotide sequences corresponding to RSQFERKR (with RKR being the first 3 amino acids of the homeodomain). For detection of the long linker variant we designed primers either corresponding to the RSQFGKCQ sequence, or GKCQERKR. We observed in control PCRs (using plasmid DNA harboring either a long or a short linker cDNA)

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that primers designed for the short linker variant still amplified the long linker one. Moreover, with linker sequences differing in only four amino acids, encoded by 12 base pars, primer pairs flanking the linker could also not be used, since both variants would be amplified in RT-qPCR experiments with almost equal efficiencies. Therefore, we used primer pairs flanking the linker region to indiscriminately amplify short and long linker variants, using non-saturating PCR (18 cycles) on total cDNA generated from total RNA. We then resolved and assessed the relative amounts of long and short linker amplicons in a second step using Fragment Analyzer (Advanced Analytical). RNA was extracted from stage 13 embryos, second instar larvae at 60 h of development, and leg or wing discs from third instar wandering larvae using the Trizol® reagent (ThermoFischer Scientific), following the manufacturer's instructions. Total RNA amounts were measured by NanoDrop and equal amounts were used to synthesize cDNA using High-Capacity RNA-to-cDNA™ Kit (ThermoFischer Scientific), following the manufacturer's instructions. Total cDNA yields were measured by NanoDrop and equal amounts were used in PCR, using in-house produced Taq polymerase. 10 ng of plasmid DNA, bearing either a long or a short transcript cDNA were used as a control. PCR product abundance was analyzed both by agarose gel electrophoresis and using Fragment Analyzer (Advanced Analytical).

Fluorescence Microscopy Imaging of live cells and FCS

Fluorescence imaging and FCS measurements were performed on two uniquely modified confocal laser scanning microscopy systems, both comprised of the ConfoCor3 system (Carl Zeiss, Jena, Germany) and consisting of either an inverted microscope for transmitted light and epifluorescence (Axiovert 200 M); a

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

VIS-laser module comprising the Ar/ArKr (458, 477, 488 and 514 nm), HeNe 543 nm and HeNe 633 nm lasers and the scanning module LSM510 META; or a Zeiss LSM780 inverted setup, comprising Diode 405 nm, Ar multiline 458, 488 and 514 nm, DPSS 561 nm and HeNe 633 nm lasers. Both instruments were modified to enable detection using silicon Avalanche Photo Detectors (SPCM-AQR-1X; PerkinElmer, USA) for imaging and FCS. Images were recorded at a 512X512 pixel resolution. C-Apochromat 40x/1.2 W UV-VIS-IR objectives were used throughout. Fluorescence intensity fluctuations were recorded in arrays of 10 consecutive measurements, each measurement lasting 10 s. Averaged curves were analyzed using the software for online data analysis or exported and fitted offline using the OriginPro 8 data analysis software (OriginLab Corporation, Northampton, MA). In either case, the nonlinear least square fitting of the autocorrelation curve was performed using the Levenberg–Marquardt algorithm. Quality of the fitting was evaluated by visual inspection and by residuals analysis. Control FCS measurements to asses the detection volume were routinely performed prior to data acquisition, using dilute solutions of known concentration of Rhodamine 6G and Alexa488 dyes. The variation between independent measurements reflects variations between cells, rather than imprecision of FCS measurements. For more details on Fluorescence Microscopy Imaging and FCS, refer to Supplement 1.

Sample size, biological and technical replicates

For the measurement of TF molecular numbers and variation (Figures 1 and 2), 7-10 larvae of each fly strain were dissected, yielding at least 15 imaginal discs, which were used in FCS analysis. For the Fkh TF, 7 pairs of salivary glands were analyzed and for AbdB, 12 genital discs were dissected from 12 larvae. More than

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

50 FCS measurements were performed in patches of neighboring cells of these dissected discs, in the regions of expression indicated in Figure 1 by arrows. Imaginal discs from the same fly strain (expressing a given endogenously-tagged TF) were analyzed on at least 3 independent instances (FCS sessions), taking place on different days (biological replicates) and for Antp, which was further analyzed in this study, more than 20 independent FCS sessions were used. As routinely done with FCS measurements in live cells, these measurements were evaluated during acquisition and subsequent analysis and, based on their quality (high counts per molecule and second, low photobleaching), were included in the calculation of concentration and variability. In Figure 2 A, *n* denotes the number of FCS measurements included in the calculations.

For experiments involving immunostainings in imaginal discs to understand the auto-regulatory behavior of Antp (Figures 3 and 5-7 and their figure supplements, except for the temporally-resolved auto-activating and repressing study of Antp in Figure 4, as discussed above), 14-20 male and female flies were mated in bottles and 10 larvae were selected by means of fluorescent balancers and processed downstream. Up to 20 imaginal discs were visualized by fluorescence microscopy and high resolution Z-stacks were acquired for 3-5 representative discs or disc regions of interest per experiment. All experiments were performed in triplicate, except for the temporal analysis of Antp auto-regulatory behavior in Figure 4, which was performed 5 times and the quantification of repression efficiency of short and long linker Antp isoforms in Figure 7, which was performed 4 times.

For the quantification of transcript variant abundance in Figure 7 K, RNA and thus cDNA was prepared from each stage 3 independent times (biological replicates) and the transcript abundance per RNA/cDNA sample was also analyzed 3 times.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For the experiment on the requirement of Antp maintenance for proper leg development in Figure 8 – figure supplement 1, A, more than 100 adult flies have been analyzed and this experiment has been performed more than 10 times independently.

References

Affolter, M., Percival-Smith, A., Muller, M., Leupin, W., and Gehring, W.J. (1990). DNA binding properties of the purified Antennapedia homeodomain. Proc Natl Acad Sci U S A *87*, 4093-4097.

Appel, B., and Sakonju, S. (1993). Cell-Type-Specific Mechanisms of Transcriptional Repression by the Homeotic Gene-Products Ubx and Abd-a in Drosophila Embryos. Embo Journal *12*, 1099-1109.

Baker, N.E., Yu, S., and Han, D. (1996). Evolution of proneural atonal expression during distinct regulatory phases in the developing Drosophila eye. Current biology : CB *6*, 1290-1301.

Becskei, A., and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. Nature *405*, 590-593.

Blake, W.J., M, K.A., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. Nature *422*, 633-637.

Boettiger, A.N., and Levine, M. (2013). Rapid transcription fosters coordinate snail expression in the Drosophila embryo. Cell reports *3*, 8-15.

Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. (2007). Isotope-labeled protein standards: toward absolute quantitative proteomics. Mol Cell Proteomics *6*, 2139-2149.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Buszczak, M., Paterno, S., Lighthouse, D., Bachman, J., Planck, J., Owen, S., Skora, A.D., Nystul, T.G., Ohlstein, B., Allen, A.*, et al.* (2007). The carnegie protein trap library: a versatile tool for Drosophila developmental studies. Genetics *175*, 1505-1531.

Calleja, M., Moreno, E., Pelaz, S., and Morata, G. (1996). Visualization of gene expression in living adult Drosophila. Science *274*, 252-255.

Clark, N.M., Hinde, E., Winter, C.M., Fisher, A.P., Crosti, G., Blilou, I., Gratton, E., Benfey, P.N., and Sozzani, R. (2016). Tracking transcription factor mobility and interaction in Arabidopsis roots with fluorescence correlation spectroscopy. eLife *5*.

Crickmore, M.A., Ranade, V., and Mann, R.S. (2009). Regulation of Ubx expression by epigenetic enhancer silencing in response to Ubx levels and genetic variation. PLoS genetics *5*, e1000633.

Dadiani, M., van Dijk, D., Segal, B., Field, Y., Ben-Artzi, G., Raveh-Sadka, T., Levo, M., Kaplow, I., Weinberger, A., and Segal, E. (2013). Two DNA-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise. Genome research *23*, 966-976.

Dublanche, Y., Michalodimitrakis, K., Kummerer, N., Foglierini, M., and Serrano, L. (2006). Noise in transcription negative feedback loops: simulation and experimental analysis. Mol Syst Biol *2*, 41.

Ejsmont, R.K., Bogdanzaliewa, M., Lipinski, K.A., and Tomancak, P. (2011). Production of fosmid genomic libraries optimized for liquid culture recombineering and cross-species transgenesis. Methods in molecular biology *772*, 423-443.

Ejsmont, R.K., Sarov, M., Winkler, S., Lipinski, K.A., and Tomancak, P. (2009). A toolkit for high-throughput, cross-species gene engineering in Drosophila. Nat Methods *6*, 435-437.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Emerald, B.S., and Cohen, S.M. (2004). Spatial and temporal regulation of the homeotic selector gene Antennapedia is required for the establishment of leg identity in Drosophila. Dev Biol *267*, 462-472.

Engstrom, Y., Schneuwly, S., and Gehring, W.J. (1992). Spatial and Temporal Expression of an Antennapedia Lac Z Gene Construct Integrated into the Endogenous Antennapedia Gene of Drosophila-Melanogaster. Dev Genes Evol *201*, 65-80.

Franz, K., Singh, A., and Weinberger, L.S. (2011). Lentiviral vectors to study stochastic noise in gene expression. Methods in enzymology *497*, 603-622.

Garaulet, D.L., Foronda, D., Calleja, M., and Sanchez-Herrero, E. (2008). Polycomb-dependent Ultrabithorax Hox gene silencing induced by high Ultrabithorax levels in Drosophila. Development *135*, 3219-3228.

Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007a). Probing the limits to positional information. Cell *130*, 153-164.

Gregor, T., Wieschaus, E.F., McGregor, A.P., Bialek, W., and Tank, D.W. (2007b). Stability and nuclear dynamics of the bicoid morphogen gradient. Cell *130*, 141-152.

Grieder, N.C., de Cuevas, M., and Spradling, A.C. (2000). The fusome organizes the microtubule network during oocyte differentiation in Drosophila. Development *127*, 4253-4264.

Gronlund, A., Lotstedt, P., and Elf, J. (2013). Transcription factor binding kinetics constrain noise suppression via negative feedback. Nat Commun *4*, 1864.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hanke, S., Besir, H., Oesterhelt, D., and Mann, M. (2008). Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. J Proteome Res *7*, 1118-1130.

Holloway, D.M., Lopes, F.J., da Fontoura Costa, L., Travencolo, B.A., Golyandina, N., Usevich, K., and Spirov, A.V. (2011). Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in Drosophila segmentation. PLoS computational biology *7*, e1001069.

Holloway, D.M., and Spirov, A.V. (2015). Mid-embryo patterning and precision in Drosophila segmentation: Kruppel dual regulation of hunchback. PLoS One *10*, e0118450.

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ramia, M., Tarone, A., Turlapati, L., Zichner, T., Zhu, D., Lyman, R.*, et al.* (2014). Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome research.

Ito, K., Awano, W., Suzuki, K., Hiromi, Y., and Yamamoto, D. (1997). The Drosophila mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. Development *124*, 761-771.

Jarman, A.P., Sun, Y., Jan, L.Y., and Jan, Y.N. (1995). Role of the proneural gene, atonal, in formation of Drosophila chordotonal organs and photoreceptors. Development *121*, 2019-2030.

Kang, M., Piliszek, A., Artus, J., and Hadjantonakis, A.K. (2013). FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. Development *140*, 267-279.

Kaur, G., Costa, M.W., Nefzger, C.M., Silva, J., Fierro-Gonzalez, J.C., Polo, J.M., Bell, T.D., and Plachta, N. (2013). Probing transcription factor diffusion dynamics in

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the living mammalian embryo with photoactivatable fluorescence correlation spectroscopy. Nat Commun *4*, 1637.

Kelso, R.J., Buszczak, M., Quinones, A.T., Castiblanco, C., Mazzalupo, S., and Cooley, L. (2004). Flytrap, a database documenting a GFP protein-trap insertion screen in Drosophila melanogaster. Nucleic Acids Res *32*, D418-420.

Kim, H.D., Shay, T., O'Shea, E.K., and Regev, A. (2009). Transcriptional regulatory circuits: predicting numbers from alphabets. Science *325*, 429-432.

Krasnow, M.A., Saffman, E.E., Kornfeld, K., and Hogness, D.S. (1989). Transcriptional activation and repression by Ultrabithorax proteins in cultured Drosophila cells. Cell *57*, 1031-1043.

Krieger, J.W., Singh, A.P., Bag, N., Garbe, C.S., Saunders, T.E., Langowski, J., and Wohland, T. (2015). Imaging fluorescence (cross-) correlation spectroscopy in live cells and organisms. Nat Protoc *10*, 1948-1974.

Kumar, R.M., Cahan, P., Shalek, A.K., Satija, R., DaleyKeyser, A.J., Li, H., Zhang, J., Pardee, K., Gennert, D., Trombetta, J.J.*, et al.* (2014). Deconstructing transcriptional heterogeneity in pluripotent stem cells. Nature *516*, 56-61.

Kuziora, M.A., and McGinnis, W. (1988). Autoregulation of a Drosophila homeotic selector gene. Cell *55*, 477-485.

Lam, C.S., Mistri, T.K., Foo, Y.H., Sudhaharan, T., Gan, H.T., Rodda, D., Lim, L.H., Chou, C., Robson, P., Wohland, T.*, et al.* (2012). DNA-dependent Oct4-Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy. The Biochemical journal *448*, 21-33.

Le, T., Liang, Z., Patel, H., Yu, M.H., Sivasubramaniam, G., Slovitt, M., Tanentzapf, G., Mohanty, N., Paul, S.M., Wu, V.M.*, et al.* (2006). A new family of Drosophila

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

balancer chromosomes with a w- dfd-GMR yellow fluorescent protein marker. Genetics *174*, 2255-2257.

Levine, M., Hafen, E., Garber, R.L., and Gehring, W.J. (1983). Spatial distribution of Antennapedia transcripts during Drosophila development. EMBO J *2*, 2037-2046.

Li, Y., Wang, F., Lee, J.A., and Gao, F.B. (2006). MicroRNA-9a ensures the precise specification of sensory organ precursors in Drosophila. Genes Dev *20*, 2793-2805.

Little, S.C., Tikhonov, M., and Gregor, T. (2013). Precise developmental gene expression arises from globally stochastic transcriptional activity. Cell *154*, 789-800.

Lo, C.A., Kays, I., Emran, F., Lin, T.J., Cvetkovska, V., and Chen, B.E. (2015). Quantification of Protein Levels in Single Living Cells. Cell reports *13*, 2634-2644.

Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes, J., Walczak, A.M., Coppey, M., and Dostatni, N. (2013). Live imaging of bicoid-dependent transcription in Drosophila embryos. Current biology : CB *23*, 2135-2139.

Mistri, T.K., Devasia, A.G., Chu, L.T., Ng, W.P., Halbritter, F., Colby, D., Martynoga, B., Tomlinson, S.R., Chambers, I., Robson, P.*, et al.* (2015). Selective influence of Sox2 on POU transcription factor binding in embryonic and neural stem cells. EMBO reports *16*, 1177-1191.

Morin, X., Daneman, R., Zavortink, M., and Chia, W. (2001). A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila. Proc Natl Acad Sci U S A *98*, 15050-15055.

Muller, C.B., Loman, A., Pacheco, V., Koberling, F., Willbold, D., Richtering, W., and Enderlein, J. (2008). Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. Epl-Europhys Lett *83*.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Muller, M., Affolter, M., Leupin, W., Otting, G., Wuthrich, K., and Gehring, W.J. (1988). Isolation and sequence-specific DNA binding of the Antennapedia homeodomain. EMBO J *7*, 4299-4304.

Nevozhay, D., Adams, R.M., Murphy, K.F., Josic, K., and Balazsi, G. (2009). Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. Proc Natl Acad Sci U S A *106*, 5123-5128.

Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840-846.

Organista, M.F., Martin, M., de Celis, J.M., Barrio, R., Lopez-Varea, A., Esteban, N., Casado, M., and de Celis, J.F. (2015). The Spalt Transcription Factors Generate the Transcriptional Landscape of the Drosophila melanogaster Wing Pouch Central Region. PLoS genetics *11*, e1005370.

Papadopoulos, D.K., Krmpot, A.J., Nikolic, S.N., Krautz, R., Terenius, L., Tomancak, P., Rigler, R., Gehring, W.J., and Vukojevic, V. (2015). Probing the kinetic landscape of Hox transcription factor-DNA binding in live cells by massively parallel Fluorescence Correlation Spectroscopy. Mechanisms of development *138 Pt 2*, 218-225.

Papadopoulos, D.K., Resendez-Perez, D., Cardenas-Chavez, D.L., Villanueva-Segura, K., Canales-del-Castillo, R., Felix, D.A., Funfschilling, R., and Gehring, W.J. (2011). Functional synthetic Antennapedia genes and the dual roles of YPWM motif and linker size in transcriptional activation and repression. Proc Natl Acad Sci U S A *108*, 11959-11964.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Papadopoulos, D.K., Skouloudaki, K., Adachi, Y., Samakovlis, C., and Gehring, W.J. (2012). Dimer formation via the homeodomain is required for function and specificity of Sex combs reduced in Drosophila. Dev Biol *367*, 78-89.

Papadopoulos, D.K., Vukojevic, V., Adachi, Y., Terenius, L., Rigler, R., and Gehring, W.J. (2010). Function and specificity of synthetic Hox transcription factors in vivo. Proc Natl Acad Sci U S A *107*, 4087-4092.

Pare, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y., and McGinnis, W. (2009). Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting. Current biology : CB *19*, 2037-2042.

Pelaez, N., Gavalda-Miralles, A., Wang, B., Navarro, H.T., Gudjonson, H., Rebay, I., Dinner, A.R., Katsaggelos, A.K., Amaral, L.A., and Carthew, R.W. (2015). Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. eLife *4*.

Perez-Camps, M., Tian, J., Chng, S.C., Sem, K.P., Sudhaharan, T., Teh, C., Wachsmuth, M., Korzh, V., Ahmed, S., and Reversade, B. (2016). Quantitative imaging reveals real-time Pou5f3-Nanog complexes driving dorsoventral mesendoderm patterning in zebrafish. eLife *5*.

Plachta, N., Bollenbach, T., Pease, S., Fraser, S.E., and Pantazis, P. (2011). Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. Nat Cell Biol *13*, 117-123.

Quinones-Coello, A.T., Petrella, L.N., Ayers, K., Melillo, A., Mazzalupo, S., Hudson, A.M., Wang, S., Castiblanco, C., Buszczak, M., Hoskins, R.A.*, et al.* (2007). Exploring strategies for protein trapping in Drosophila. Genetics *175*, 1089-1104.

Raj, A., and van Oudenaarden, A. (2009). Single-molecule approaches to stochastic gene expression. Annual review of biophysics *38*, 255-270.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ryoo, H.D., and Mann, R.S. (1999). The control of trunk Hox specificity and activity by Extradenticle. Genes Dev *13*, 1704-1716.

Saffman, E.E., and Krasnow, M.A. (1994). A differential response element for the homeotics at the Antennapedia P1 promoter of Drosophila. Proc Natl Acad Sci U S A *91*, 7420-7424.

Sanchez, A., Garcia, H.G., Jones, D., Phillips, R., and Kondev, J. (2011). Effect of promoter architecture on the cell-to-cell variability in gene expression. PLoS computational biology *7*, e1001100.

Sarov, M., Barz, C., Jambor, H., Hein, M.Y., Schmied, C., Suchold, D., Stender, B., Janosch, S., K, J.V., Krishnan, R.T.*, et al.* (2016). A genome-wide resource for the analysis of protein localisation in Drosophila. eLife *5*, e12068.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.*, et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676-682.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Sheth, R., Bastida, M.F., Kmita, M., and Ros, M. (2014). "Self-Regulation," A New Facet of Hox Genes' Function. Dev Dynam *243*, 182-191.

Shimoga, V., White, J.T., Li, Y., Sontag, E., and Bleris, L. (2013). Synthetic mammalian transgene negative autoregulation. Mol Syst Biol *9*, 670.

Slattery, M., Voutev, R., Ma, L., Negre, N., White, K.P., and Mann, R.S. (2013). Divergent transcriptional regulatory logic at the intersection of tissue growth and developmental patterning. PLoS genetics *9*, e1003753.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Smolander, O.P., Kandhavelu, M., Mannerstrom, H., Lihavainen, E., Kalaichelvan, S., Healy, S., Yli-Harja, O., Karp, M., and Ribeiro, A.S. (2011). Cell-to-cell diversity in protein levels of a gene driven by a tetracycline inducible promoter. BMC molecular biology *12*, 21.

Stroeher, V.L., Gaiser, J.C., and Garber, R.L. (1988). Alternative RNA splicing that is spatially regulated: generation of transcripts from the Antennapedia gene of Drosophila melanogaster with different protein-coding regions. Molecular and cellular biology *8*, 4143-4154.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472-474.

Swain, P.S., Elowitz, M.B., and Siggia, E.D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci U S A *99*, 12795-12800.

Szaloki, N., Krieger, J.W., Komaromi, I., Toth, K., and Vamosi, G. (2015). Evidence for Homodimerization of the c-Fos Transcription Factor in Live Cells Revealed by Fluorescence Microscopy and Computer Modeling. Molecular and cellular biology *35*, 3785-3798.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533-538.

Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A *98*, 8614-8619.

Tiwari, M., Mikuni, S., Muto, H., and Kinjo, M. (2013). Determination of dissociation constant of the NFkappaB p50/p65 heterodimer using fluorescence cross-correlation

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

spectroscopy in the living cell. Biochemical and biophysical research communications *436*, 430-435.

Tsutsumi, M., Muto, H., Myoba, S., Kimoto, M., Kitamura, A., Kamiya, M., Kikukawa, T., Takiya, S., Demura, M., Kawano, K.*, et al.* (2016). In vivo fluorescence correlation spectroscopy analyses of FMBP-1, a silkworm transcription factor. FEBS open bio *6*, 106-125.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat Rev Genet *10*, 252-263.

Venken, K.J., Schulze, K.L., Haelterman, N.A., Pan, H., He, Y., Evans-Holm, M., Carlson, J.W., Levis, R.W., Spradling, A.C., Hoskins, R.A.*, et al.* (2011). MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat Methods *8*, 737-743.

Vukojevic, V., Heidkamp, M., Ming, Y., Johansson, B., Terenius, L., and Rigler, R. (2008). Quantitative single-molecule imaging by confocal laser scanning microscopy. Proc Natl Acad Sci U S A *105*, 18176-18181.

Vukojevic, V., Papadopoulos, D.K., Terenius, L., Gehring, W.J., and Rigler, R. (2010). Quantitative study of synthetic Hox transcription factor-DNA interactions in live cells. Proc Natl Acad Sci U S A *107*, 4093-4098.

Winslow, G.M., Hayashi, S., Krasnow, M., Hogness, D.S., and Scott, M.P. (1989). Transcriptional activation by the Antennapedia and fushi tarazu proteins in cultured Drosophila cells. Cell *57*, 1017-1030.

Wirz, J., Fessler, L.I., and Gehring, W.J. (1986). Localization of the Antennapedia protein in Drosophila embryos and imaginal discs. EMBO J *5*, 3327-3334.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Zink, B., Engstrom, Y., Gehring, W.J., and Paro, R. (1991). Direct interaction of the Polycomb protein with Antennapedia regulatory sequences in polytene chromosomes of Drosophila melanogaster. EMBO J *10*, 153-162.

Supplemental information

Supplement 1: Background on Fluorescence Microscopy Imaging and FCS

Two individually modified instruments (Zeiss, LSM 510 and 780, ConfoCor 3) with fully integrated FCS/CLSM optical pathways were used for imaging. The detection efficiency of CLSM imaging was significantly improved by the introduction of APD detectors. As compared to PMTs, which are normally used as detectors in conventional CLSM, the APDs are characterized by higher quantum yield and collection efficiency – about 70 % in APDs as compared to 15 – 25 % in PMTs, higher gain, negligible dark current and better efficiency in the red part of the spectrum. Enhanced fluorescence detection efficiency enabled image collection using fast scanning ($1 - 5\ \mu s/pixel$). This enhances further the signal-to-noise-ratio by avoiding fluorescence loss due to triplet state formation, enabling fluorescence imaging with single-molecule sensitivity. In addition, low laser intensities ($150 - 750\ \mu W$) could be applied for imaging, significantly reducing the photo-toxicity (Vukojevic et al., 2008).

FCS measurements are performed by recording fluorescence intensity fluctuations in a very small, approximately ellipsoidal observation volume element (OVE) (about $0.2 \mu m$ wide and $1 \mu m$ long) that is generated in imaginal disc cells by focusing the laser light through the microscope objective and by collecting the fluorescence light through the same objective using a pinhole in front of the detector

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to block out-of-focus light. The fluorescence intensity fluctuations, caused by fluorescently labeled molecules passing through the OVE are analyzed using temporal autocorrelation analysis.

In temporal autocorrelation analysis we first derive the autocorrelation function $G(\tau)$:

$$G(\tau) = 1 + \frac{\langle \delta I(t) \cdot \delta I(t+\tau) \rangle}{\langle I(t) \rangle^2},$$

where $\delta I(t) = I(t) - \langle I(t) \rangle$ is the deviation from the mean intensity at time t and $\delta I(t+\tau) = I(t+\tau) - \langle I(t) \rangle$ is the deviation from the mean intensity at time $t + \tau$. For further analysis, an autocorrelation curve is derived by plotting $G(\tau)$ as a function of the lag time, i.e. the autocorrelation time τ.

To derive information about molecular numbers and their corresponding diffusion time, the experimentally obtained autocorrelation curves are compared to autocorrelation functions derived for different model systems. A model describing free three dimensional (3D) diffusion of two components and triplet formation was used in this study:

$$G(\tau) = 1 + \frac{1}{N}\left(\frac{1-y}{\left(1+\frac{\tau}{\tau_{D_1}}\right) \cdot \sqrt{1+\frac{w_{xy}^2 \tau}{w_z^2 \tau_{D_1}}}} + \frac{y}{\left(1+\frac{\tau}{\tau_{D_2}}\right) \cdot \sqrt{1+\frac{w_{xy}^2 \tau}{w_z^2 \tau_{D_2}}}} \right) \cdot \left(1 + \frac{T}{1-T} \cdot e^{-\frac{\tau}{\tau_T}}\right)$$

In the above equation, *N* is the average number of molecules in the OVE; *y* is the fraction of the slowly moving Antp-eGFP molecules; τ_{D_1} is the diffusion time of the free Antp-eGFP molecules; τ_{D_2} is the diffusion time of Antp-eGFP molecules undergoing interactions with the DNA; w_{xy} and w_z are radial and axial parameters, respectively, related to spatial properties of the OVE; T is the average equilibrium fraction of molecules in the triplet state; and τ_T the triplet correlation time related to rate constants for intersystem crossing and the triplet decay.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Spatial properties of the detection volume, represented by the square of the ratio of the axial and radial parameters ($\left(\frac{w_z}{w_{xy}}\right)^2$), are determined in calibration measurements performed using a solution of Rhodamine 6G for which the diffusion coefficient (D) is known to be $D_{Rh6G} = 4.1 \cdot 10^{-10}\, m^2 s^{-1}$ (Muller et al., 2008). The diffusion time, τ_D, measured by FCS, is related to the translation diffusion coefficient *D* by:

$$\tau_D = \frac{w_{xy}^2}{4D}.$$

<u>Supplement 2: Calculation of the concentration of endogenous TFs and average number of molecules in imaginal disc cell nuclei from FCS measurements (exemplified for Antp)</u>

The Observation Volume Element (OVE), which can be approximated by a prolate ellipsoid, has a volume of:

$$V_{OVE} = \pi^{\frac{3}{2}} \cdot w_{xy}^2 \cdot z_0 = 5.57 \cdot 0.1847^{\mathbf{2}} \cdot 1 = 0.223 \cdot 10^{-18}\, m^3 = 0.19 \cdot 10^{-15} L$$

Therefore, one fluorescent molecule in the OVE yields equal concentration to: $\frac{1}{0.19 \cdot 10^{-15}} = 5.263 \cdot 10^{15}$ molecules per L.

$1\, M$ of Antp-eGFP molecules equals $6.022 \cdot 10^{23}$ molecules/L,

$x\, M$ of Antp-eGFP molecules equal $5.263 \cdot 10^{15}$ molecules/L.

Therefore:

$$x = \frac{5.263 \cdot 10^{15}}{6.022 \cdot 10^{23}} = 0.874 \cdot 10^{-8}\, M = 8.74\, nM$$

This indicates that one molecule in the OVE corresponds to $8.74\, nM$ concentration.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The wing disc cells within the Antp expression domain (prescutum precursors) are not spherical, but rather ellipsoidal. Their axes were determined by fluorescence imaging to be 1.4 μm in the transverse dimension and 2.8 μm in the longitudinal. The volume of the nucleus was approximated by the volume of a prolate ellipsoid:

$$V_{nucleus} = \frac{4}{3}\pi a^2 b = \frac{4}{3} \cdot 3.14 \cdot (1.4 \cdot 10^{-6})^2 \cdot 2.8 \cdot 10^{-6}\, m^3 = 22.99 \cdot 10^{-18}\, m^3$$

$$= 22.99 \cdot 10^{-15}\, L$$

Therefore, the OVE represents roughly $1/103$ of the nuclear volume:

$$\frac{V_{nucleus}}{V_{OVE}} = 121$$

and the number of molecules in Antp-eGFP nuclei is on the average $57.37 \cdot 121 \approx 6942$ molecules.

The concentration of Antp in the wing disc nuclei is calculated as follows:

$1\, mol$ of Antp-eGFP molecules equals $6.023 \cdot 10^{23}$,

$n\, mol$ of Antp-eGFP molecules equal 6942.

Thus $n_{Antp-eGFP} = \frac{6942}{6.022 \cdot 10^{23}} = 1152.77 \cdot 10^{-23}\, mol$

Thus the concentration of Antp-eGFP within the diploid wing disc nucleus will be:

$$C_{Antp-eGFP} = \frac{1182.77 \cdot 10^{-23}}{22.99 \cdot 10^{-15}} = 50.1 \cdot 10^{-8}\, M \approx 501\, nM$$

Fitting of autocorrelation curves using a model for free three dimensional (3D) diffusion of two components and triplet, revealed that on the average 63% of the total Antp-eGFP molecules are fast moving and appear to be freely diffusing in the nucleus, whereas 37% are slow and are likely participating in the formation of complexes with the DNA.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplement 3: Calculation of the endogenous apparent dissociation constant of Antp from FCS measurements on third instar wing and leg imaginal discs

The endogenously tagged Antp-eGFP TF undergoes both specific and non-specific interactions with DNA. Assuming that non-specific interactions precede the specific ones, a two-step process of consecutive reactions is anticipated (Vukojevic et al., 2010):

$$DNA_{ns} + Antp - eGFP \underset{k_{-ns}}{\overset{k_{ns}}{\Leftrightarrow}} (DNA - Antp - eGFP)_{ns}$$
$$(DNA - Antp - eGFP)_{ns} + DNA_s \underset{k_{-s}}{\overset{k_s}{\Leftrightarrow}} (DNA - Antp - eGFP)_s \quad \text{(S1)}$$

The turnover rate for the non-specific complex is:

$$\frac{d[(DNA-Antp-eGFP)_{ns}]}{dt} = k_{ns} \cdot [DNA_{ns}] \cdot [Antp - eGFP] - (k_{-ns} + k_s \cdot [DNA_s]) \cdot [(DNA - Antp - eGFP)_{ns}] + k_{-s} \cdot [(DNA - Antp - eGFP)_s] \quad \text{(S2)}$$

Assuming a quasi-steady state approximation:

$$\frac{d[(DNA-Antp-eGFP)_{ns}]}{dt} = 0 \quad \text{(S3)}$$

$$(k_{-ns} + k_s \cdot [DNA_s]) \cdot [(DNA - Antp - eGFP)_{ns}] = k_{ns} \cdot [DNA_{ns}] \cdot [Antp - eGFP] + k_{-s} \cdot [(DNA - Antp - eGFP)_s] \quad \text{(S4)}$$

Using the mass balance equation to express the concentration of the free TF:

$$[Antp - eGFP] = [Antp - eGFP]_0 - [(DNA - Antp - eGFP)_{ns}] - [(DNA - Antp - eGFP)_s] \quad \text{(S5)}$$

and assuming that:

$$[DNA]_{ns} \approx [DNA]_0 \quad \text{(S6)},$$

equation (S4) becomes:

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$(k_{-ns} + k_s \cdot [DNA_s]) \cdot [(DNA - Antp - eGFP)_{ns}] = k_{ns} \cdot [DNA]_0 \cdot [Antp - eGFP]_0 - [(DNA - Antp - eGFP)_{ns} - [(DNA - Antp - eGFP)_s] + k_{-s} \cdot [(DNA - Antp - eGFP)_s] \quad \text{(S7)},$$

$$(k_{-ns} + k_s \cdot [DNA_s] + k_{ns} \cdot [DNA]_0) \cdot [(DNA - Antp - eGFP)_{ns}] = k_{ns} \cdot [DNA]_0 \cdot ([Antp - eGFP]_0 - [(DNA - Antp - eGFP)_s]) + k_{-s} \cdot [(DNA - Antp - eGFP)_s] \quad \text{(S8)},$$

$$(k_{-ns} + k_s \cdot [DNA_s] + k_{ns} \cdot [DNA]_0) \cdot [(DNA - Antp - eGFP)_{ns}] = k_{ns} \cdot [DNA]_0 \cdot [Antp - eGFP]_0 - (k_{ns} \cdot [DNA]_0 - k_{-s}) \cdot [(DNA - Antp - eGFP)_s] \quad \text{(S9)},$$

$$[(DNA - Antp - eGFP)_{ns}] = \frac{k_{ns} \cdot [DNA]_0}{k_{-ns} + k_s \cdot [DNA_s] + k_{ns} \cdot [DNA]_0} \cdot [Antp - eGFP]_0 - \frac{k_{ns} \cdot [DNA]_0 - k_{-s}}{k_{-ns} + k_s \cdot [DNA_s] + k_{ns} \cdot [DNA]_0} \cdot [(DNA - Antp - eGFP)_s] \quad \text{(S10)}$$

According to equation (S10) and the FCS data presented in Figure 6 M, the slope of the linear dependence gives:

$$\frac{k_{ns} \cdot [DNA]_0}{k_{-ns} + k_s \cdot [DNA_s] + k_{ns} \cdot [DNA]_0} = 0.38 \quad \text{(S11)}$$

and the intercept:

$$\frac{k_{ns} \cdot [DNA]_0 - k_{-s}}{k_{-ns} + k_s \cdot [DNA_s] + k_{ns} \cdot [DNA]_0} \cdot [(DNA - Antp - eGFP)_s] = 5.78\, nM \quad \text{(S12)}$$

If k_{-s} is small compared to $k_{ns} \cdot [DNA]_0$ and can therefore be neglected, then:

$$0.38 \cdot [(DNA - Antp - eGFP)_s] = 5.78\, nM \quad \text{(S13)}$$

Thus, the concentration of specific complex between Antp-eGFP and DNA in the wing disc cell nuclei can be estimated to be:

$$[(DNA - Antp - eGFP)_s] = 15.26\, nM \quad \text{(S14)}$$

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The average concentration of free-diffusing Antp-eGFP molecules is determined as follows:

$$[Antp-eGFP]_{free} = [Antp-eGFP]_0 - [(DNA-Antp-eGFP)_{ns}] - [(DNA-Antp-eGFP)_s] = [Antp-eGFP]_0 - (0.38 \cdot [Antp-eGFP]_0 - 5.78) - [(DNA-Antp-eGFP)_s] = 501 - 0.38 \cdot 501 + 5.78 - 15.26 = 301.1nM$$

Using the experimentally determined concentration of specific DNA–Antp-eGFP complexes (eq. S14) and the estimated total concentration of specific DNA binding sites derived in Supplement 6, $[DNA]_s = 16.16\,nM$, we could estimate the dissociation constant for the specific DNA–Antp-eGFP complex to be:

$$K_{d,Antp}^{3rd\ instar} = \frac{[DNA_s]_{free} \cdot [Antp-eGFP]_{free}}{[(DNA-Antp-eGFP)_s]} = \frac{0.9 \cdot 301.1}{15.26} \approx 18\,nM$$

<u>Supplement 4: Approximation of the concentration of DNA binding sites in imaginal disc cell nuclei to which Antp can bind specifically</u>

The mean size of the Drosophila haploid genome is $1.75 \cdot 10^8$ bp long (Huang et al., 2014); this means that the diploid genome size is: $2 \cdot 1.75 \cdot 10^8 = 3.5 \cdot 10^8$bp long.

For specific binding to occur, usually a larger sequence is necessary. The statistical occurrence of the binding site correlates with the size of the site, rather than its sequence. For most Hox TF-Extradenticle (Exd) complexes, sequences consisting of a minimum of 10 bp have been identified, bearing the Exd consensus binding sequence (AGAT) (Ryoo and Mann, 1999). Assuming thus a minimum requirement of 10 bp to confer Antp binding specificity, this sequence may occur once every $4^{10} = 1048576 \approx 1.048 \cdot 10^6$ bp in the genome.

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

This results in $\frac{3.5 \cdot 10^8}{1.048 \cdot 10^6} = 334$ sites within a diploid nucleus of a wing imaginal disc cell.

$1\ mol$ of potential sites contains $6.022 \cdot 10^{23}$ sites

$n\ mol$ of potential sites contain 334 sites, thus:

$$n = \frac{334}{6.022 \cdot 10^{23}} = 55.46 \cdot 10^{-23}\ mol$$

However, the Drosophila heterochromatin comprises about 33% of the total genome sequence. Therefore, only 67% of the total DNA length can be considered as candidate sequences for TF-DNA interactions. This results in:

$0.67 \cdot 55.46 \cdot 10^{-23} = 37.158 \cdot 10^{-23}\ mol$ of binding sites.

Thus the concentration of specific binding sites within the diploid nucleus is:

$$C_{sites}^{specific} = \frac{n}{V_{nucleus}} = \frac{37.158 \cdot 10^{-23}}{22.99 \cdot 10^{-15}} = 1.616 \cdot 10^{-8} M = 16.16\ nM$$

Supplement 5: Calculation of the apparent Antp dissociation constant from FCS measurements in endogenously-expressing and overexpressing nuclei

We have calculated similar values of the apparent Antp-eGFP dissociation constant to the value obtained by FCS measurements in third instar imaginal discs also when the dissociation constant was calculated using the binding curve obtained by plotting all FCS measurements on Antp: third instar wing and leg imaginal discs, as well as overexpression in the antennal disc by all Gal4 drivers tested: *Dll*-Gal4 (MD23 and MD713), *ptc*-Gal4 and *69B*-Gal4 (Figure 6 N). We calculated a slope of 0.46 and an intercept of 7.29, yielding concentration of the Antp-eGFP specific complex:

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$[(DNA - Antp - eGFP)_s] = 15.92nM$$

With an average total concentration of Antp-eGFP of $[Antp - eGFP]_0 = 1849.12nM$, the average free-diffusing Antp-eGFP TF was determined similarly in this case to be:

$$[Antp - eGFP]_{free} = 993.59nM$$

We, therefore calculated a dissociation constant of:

$K_{d,Antp}^{endog.+o/e} = \frac{[DNA_s]_{free} \cdot [Antp-eGFP]_{free}}{[(DNA-Antp-eGFP)_s]} = \frac{0.24 \cdot 993.59}{15.92} = 14.98 \approx 15nM$, very similar to the value calculated with the binding data of homozygous, endogenous Antp-eGFP in third instar imaginal discs.

<u>Supplement 6: Calculation of the endogenous apparent Antp dissociation constant from FCS measurements on second instar wing and leg imaginal discs</u>

We determined the endogenous apparent dissociation constant for Antp in second instar wing and leg imaginal discs, using FCS measurements. We have calculated a slope of 0.44 and an intercept of 2.6, yielding concentration of the Antp-eGFP specific complex:

$$[(DNA - Antp - eGFP)_s] = 5.97nM$$

With an average total concentration in second instar wing and leg disc nuclei, Antp-eGFP of $[Antp - eGFP]_0 = 136.05nM$, the average free-diffusing Antp-eGFP TF was determined similarly in this case to be:

$$[Antp - eGFP]_{free} = 76.73nM$$

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We, therefore calculated a dissociation constant of:

$K_{d,Antp}^{2nd\ instar} = \frac{[DNA_s]_{free} \cdot [Antp-eGFP]_{free}}{[(DNA-Antp-eGFP)_s]} = \frac{10.19 \cdot 76.73}{5.97} = 130.97 \approx 131nM,$ different by roughly one order of magnitude from the value obtained from FCS measurements in third instar discs, expressing Antp endogenously or the sum of data on third instar endogenous expression and overexpression.

Supplement 7: Calculation of the apparent Antp dissociation constant for short and long linker Antp isoforms from FCS measurements on ectopically expressed Antp

We determined the apparent dissociation constant for Antp short and long linker isoforms in discs expressing Antp close to endogenous levels, using FCS measurements.

a) For the short linker Antp isoform, we have calculated a slope of 0.34 and an intercept of 5.31, yielding concentration of the Antp-eGFP specific complex:

$$[(DNA - Antp - eGFP)_s] = 15.62nM$$

With an average total concentration in second instar wing and leg disc nuclei, Antp-eGFP of $[Antp - eGFP]_0 = 785.28nM$, the average free-diffusing Antp-eGFP TF was determined similarly in this case to be:

$$[Antp - eGFP]_{free} = 507.97nM$$

We, therefore calculated a dissociation constant of:

bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$K_{d,Antp}^{short\ linker\ isof.,o/e} = \frac{[DNA_s]_{free} \cdot [Antp-eGFP]_{free}}{[(DNA-Antp-eGFP)_s]} = \frac{0.54 \cdot 507.97}{15.62} = 17.56 \approx 18nM$, equal to the value calculated with the binding data of homozygous, endogenous Antp-eGFP in third instar imaginal discs.

b) For the long linker Antp isoform, we have calculated a slope of 0.24 and an intercept of 3.28, yielding concentration of the Antp-eGFP specific complex:

$$[(DNA - Antp - eGFP)_s] = 13.67nM$$

With an average total concentration in second instar wing and leg disc nuclei, Antp-eGFP of $[Antp - eGFP]_0 = 1382.95nM$, the average free-diffusing Antp-eGFP TF was determined similarly in this case to be:

$$[Antp - eGFP]_{free} = 1040.65nM$$

We, therefore calculated a dissociation constant of:

$K_{d,Antp}^{long\ linker\ isof.,o/e} = \frac{[DNA_s]_{free} \cdot [Antp-eGFP]_{free}}{[(DNA-Antp-eGFP)_s]} = \frac{2.49 \cdot 1040.65}{13.67} = 189.56 \approx 190nM$, different by roughly one order of magnitude from the value obtained from FCS measurements in discs expressing the Antp short linker isoform. However, this value was found to be similar to the value obtained from FCS measurements on second instar leg and wing discs.

Figure 1

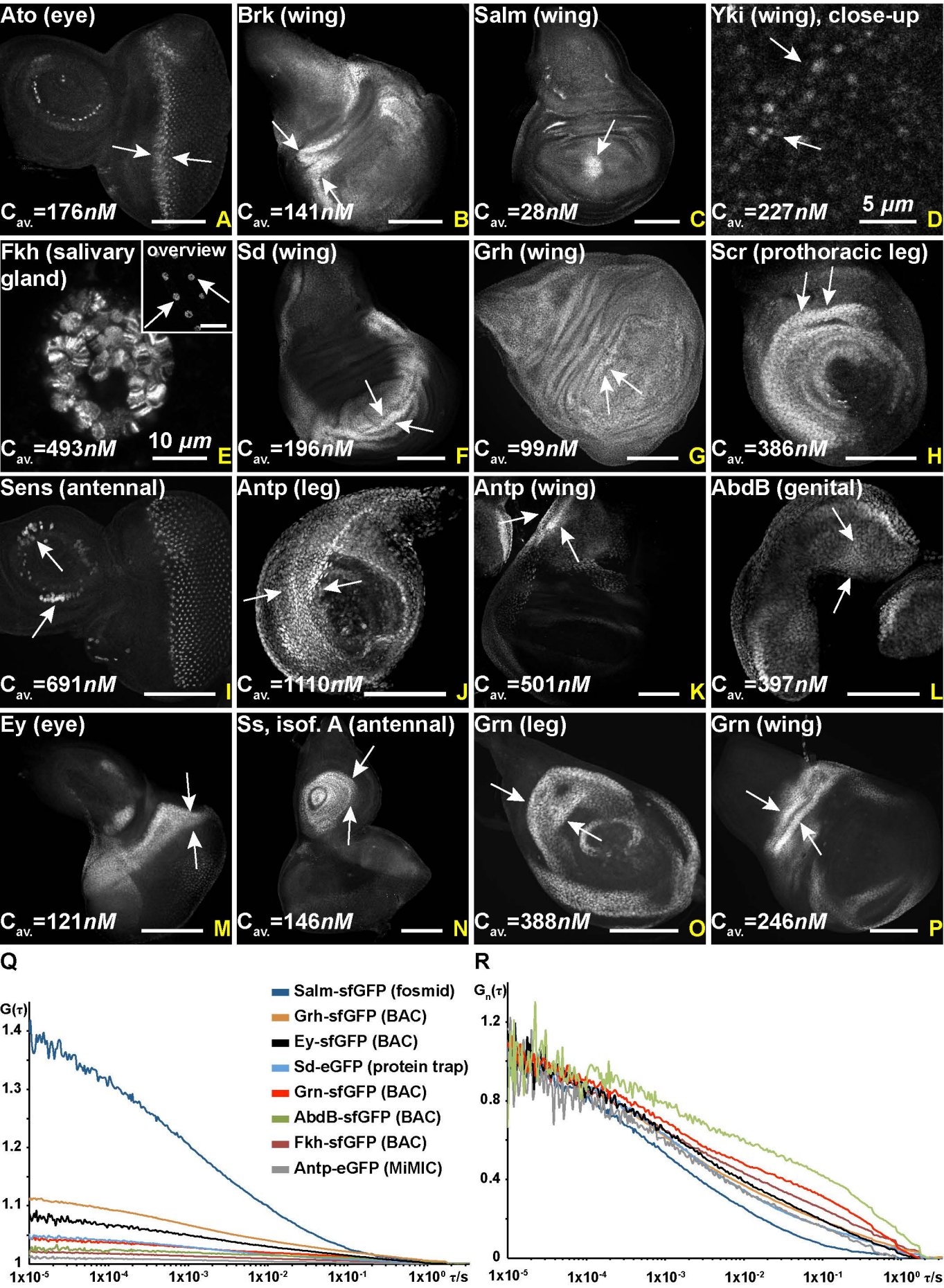

Ato (eye)
C av.=176nM
A
Brk (wing)
C av.=141nM
B
Salm (wing)
C av.=28nM
C
Yki (wing), close-up
C av.=227nM
5 µm
D
Fkh (salivary gland)
overview
C av.=493nM
10 µm
E
Sd (wing)
C av.=196nM
F
Grh (wing)
C av.=99nM
G
Scr (prothoracic leg)
C av.=386nM
H
Sens (antennal)
C av.=691nM
I
Antp (leg)
C av.=1110nM
J
Antp (wing)
C av.=501nM
K
AbdB (genital)
C av.=397nM
L
Ey (eye)
C av.=121nM
M
Ss, isof. A (antennal)
C av.=146nM
N
Grn (leg)
C av.=388nM
O
Grn (wing)
C av.=246nM
P
Q
G(τ)
1.4
1.3
1.2
1.1
1
1x10-5
1x10-4
1x10-3
1x10-2
1x10-1
1x100 τ/s
Salm-sfGFP (fosmid)
Grh-sfGFP (BAC)
Ey-sfGFP (BAC)
Sd-eGFP (protein trap)
Grn-sfGFP (BAC)
AbdB-sfGFP (BAC)
Fkh-sfGFP (BAC)
Antp-eGFP (MiMIC)
R
Gn(τ)
1.2
0.8
0.4
0
1x10-5
1x10-4
1x10-3
1x10-2
1x10-1
1x100 τ/s

Figure 2

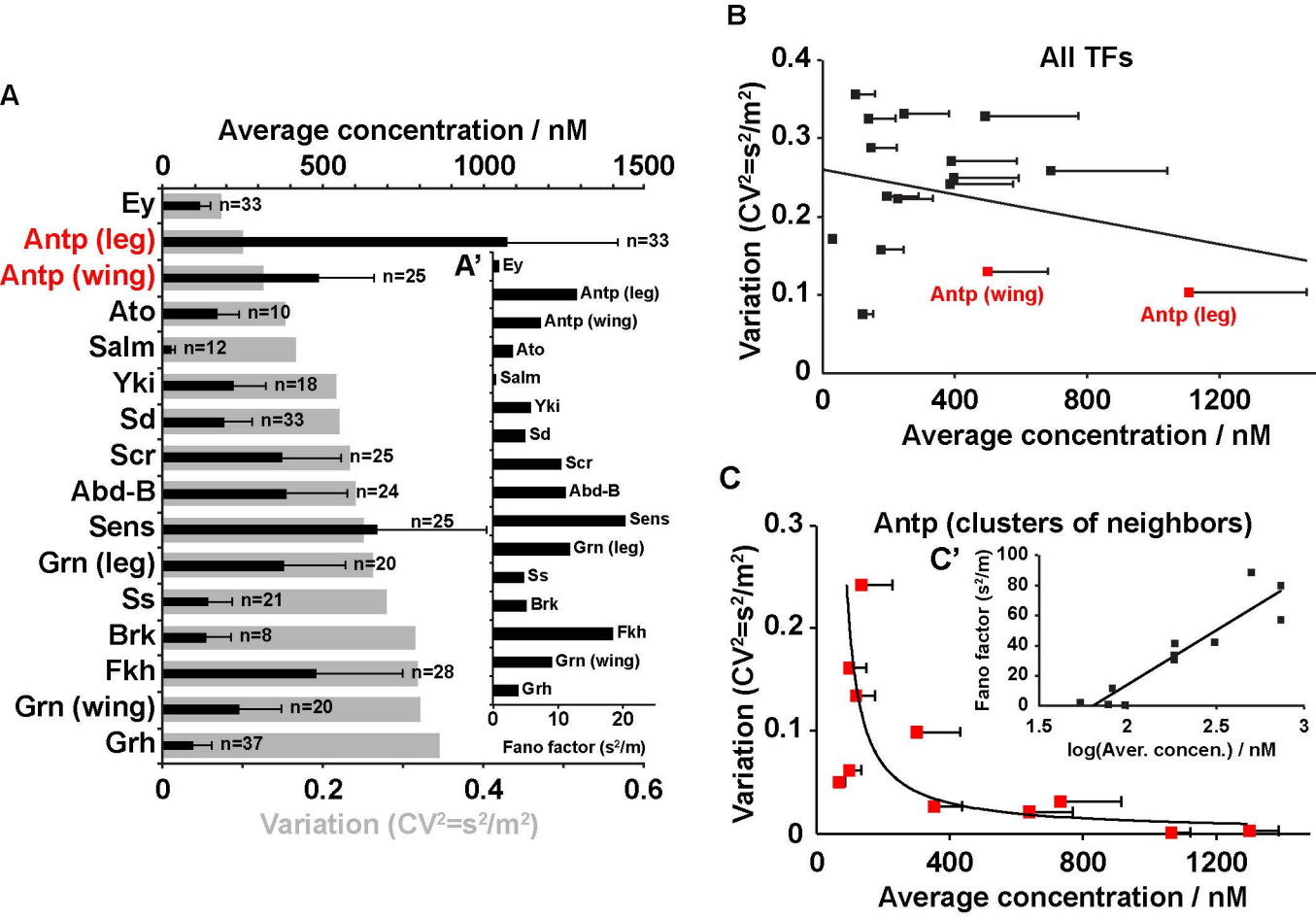

A
Average concentration / nM
0
500
1000
1500
Ey
Antp (leg)
Antp (wing)
Ato
Salm
Yki
Sd
Scr
Abd-B
Sens
Grn (leg)
Ss
Brk
Fkh
Grn (wing)
Grh
n=33
n=33
n=25
n=10
n=12
n=18
n=33
n=25
n=24
n=25
n=20
n=21
n=8
n=28
n=20
n=37
0
0.2
0.4
0.6
Variation (CV²=s²/m²)
A'
Ey
Antp (leg)
Antp (wing)
Ato
Salm
Yki
Sd
Scr
Abd-B
Sens
Grn (leg)
Ss
Brk
Fkh
Grn (wing)
Grh
0
10
20
Fano factor (s²/m)
B
All TFs
Variation (CV²=s²/m²)
0.4
0.3
0.2
0.1
0
Antp (wing)
Antp (leg)
0
400
800
1200
Average concentration / nM
C
Antp (clusters of neighbors)
Variation (CV²=s²/m²)
0.3
0.2
0.1
0
C'
Fano factor (s²/m)
100
80
60
40
20
0
1.5
2
2.5
3
log(Aver. concen.) / nM
0
400
800
1200
Average concentration / nM

Figure 3

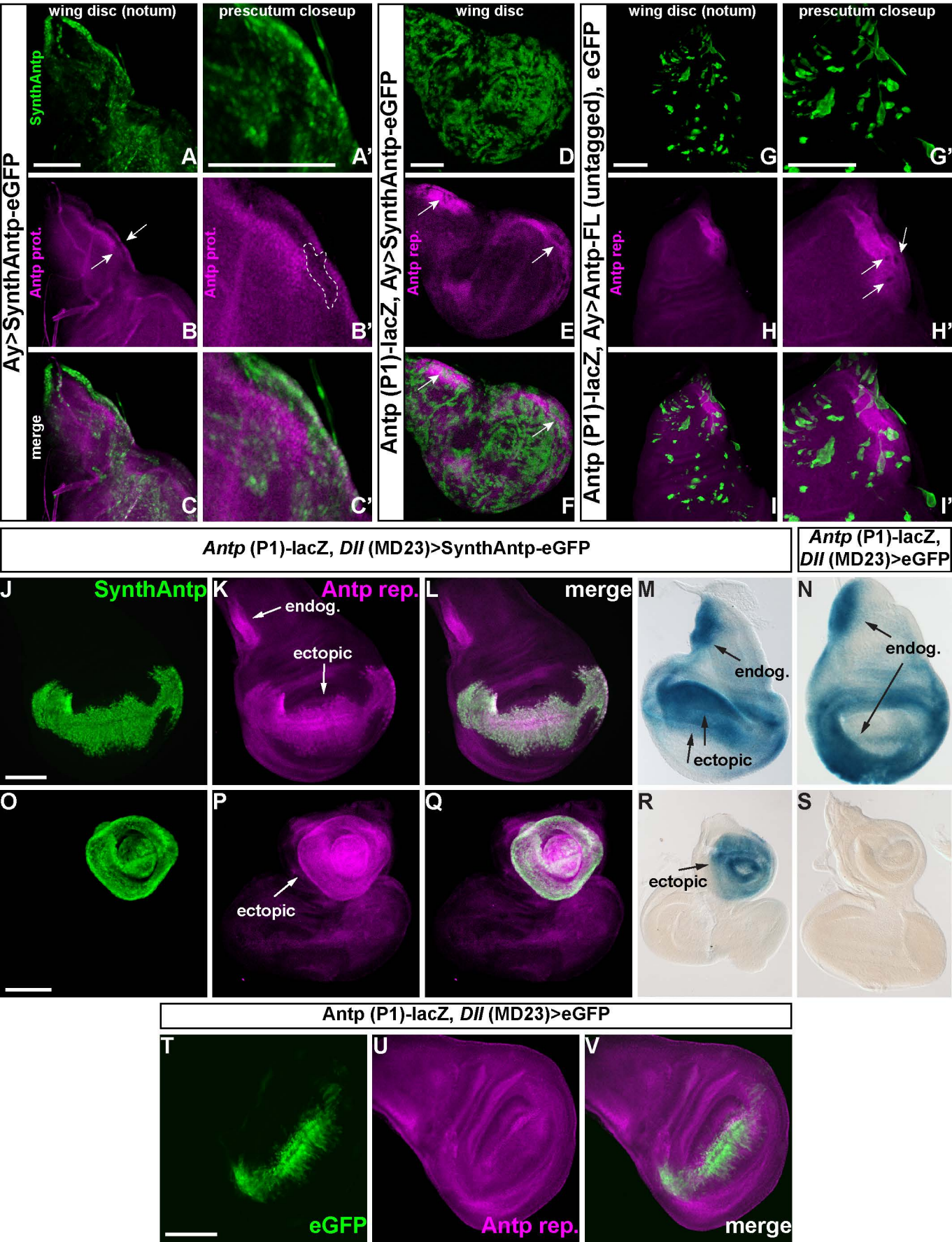
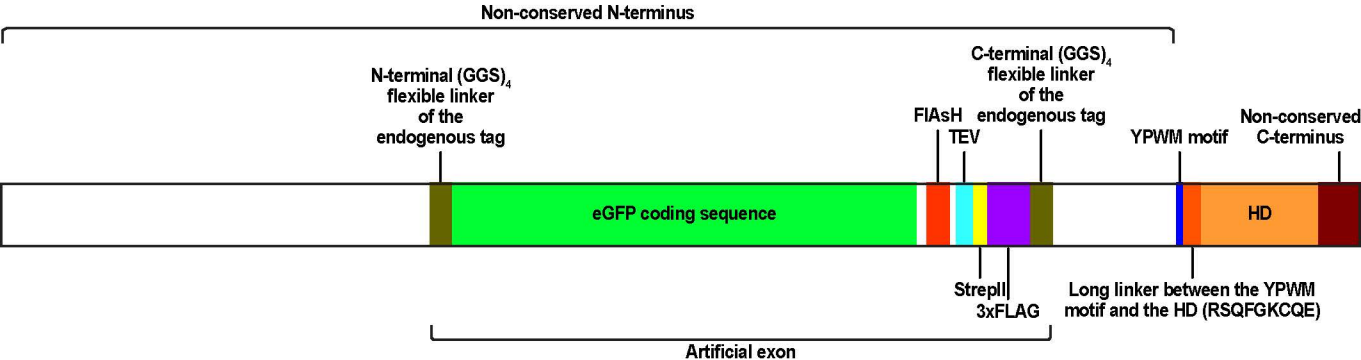
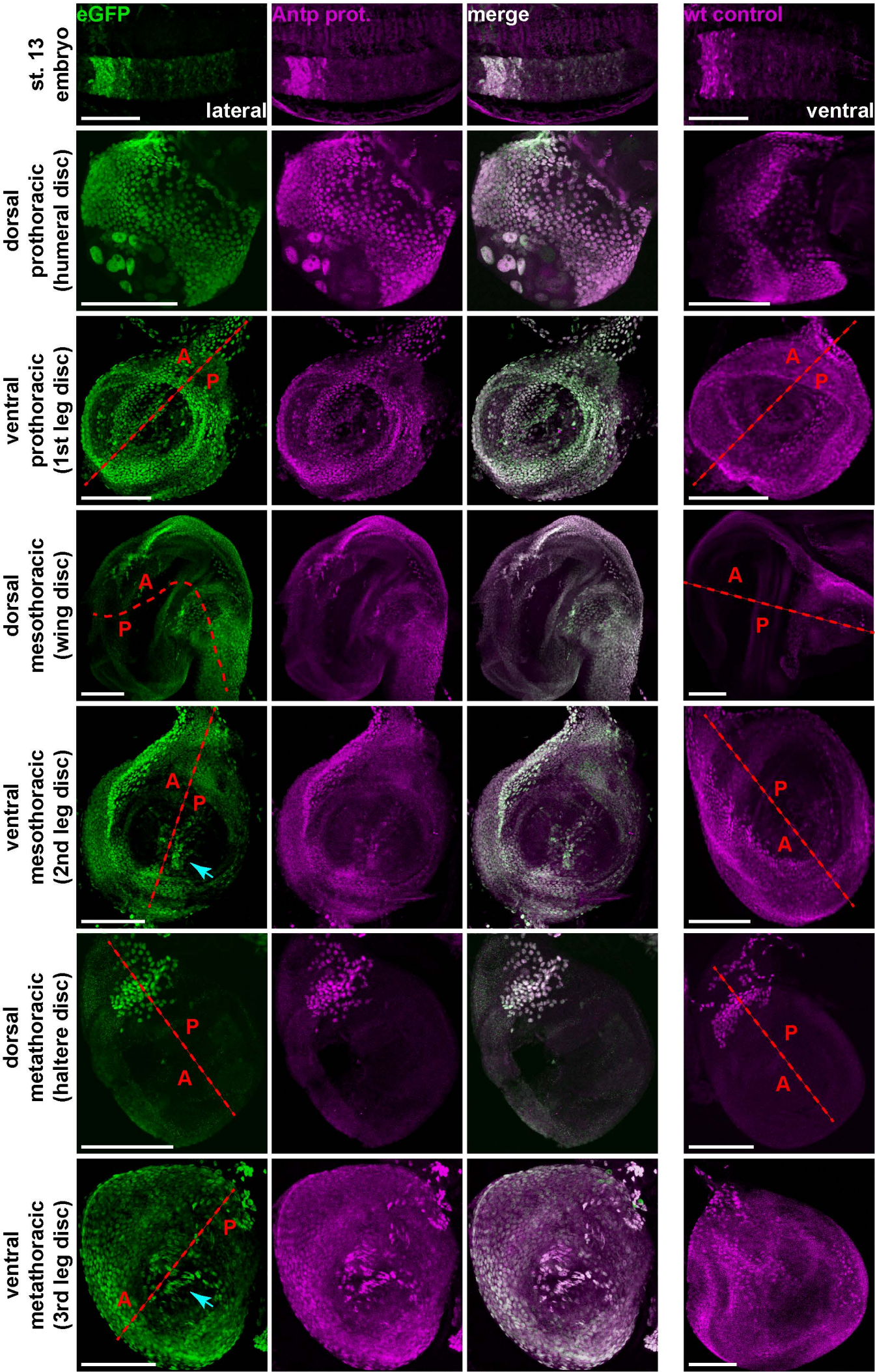

Ay>SynthAntp-eGFP
wing disc (notum)
prescutum closeup
SynthAntp
Antp prot.
merge
A
A'
B
B'
C
C'
Antp (P1)-lacZ, Ay>SynthAntp-eGFP
wing disc
Antp rep.
D
E
F
Antp (P1)-lacZ, Ay>Antp-FL (untagged), eGFP
wing disc (notum)
prescutum closeup
Antp rep.
G
G'
H
H'
I
I'
Antp (P1)-lacZ, Dll (MD23)>SynthAntp-eGFP
Antp (P1)-lacZ, Dll (MD23)>eGFP
J
SynthAntp
K
Antp rep.
endog.
ectopic
L
merge
M
endog.
ectopic
N
endog.
O
P
ectopic
Q
R
ectopic
S
Antp (P1)-lacZ, Dll (MD23)>eGFP
T
eGFP
U
Antp rep.
V
merge

Figure 3 - figure supplement 1



bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is available under aCC-BY-NC-ND 4.0 International license.

Figure 3 - figure supplement 2



bioRxiv preprint doi: https://doi.org/10.1101/115220; this version posted March 14, 2017. The copyright holder for this preprint (wh
Not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is m
available under aCC-BY-NC-ND 4.0 International license.

Figure 3 - figure supplement 3

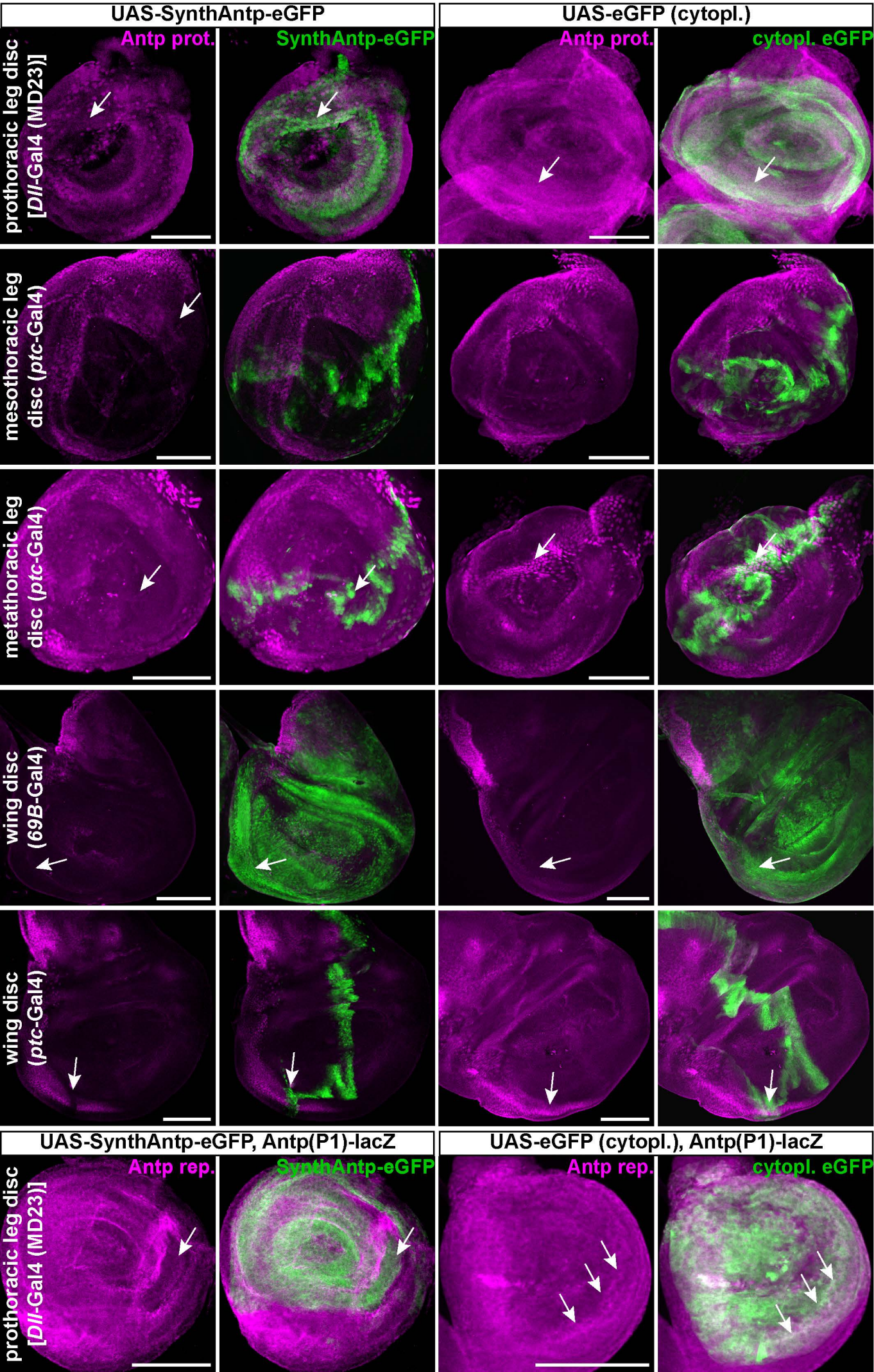


Figure 3 - figure supplement 4

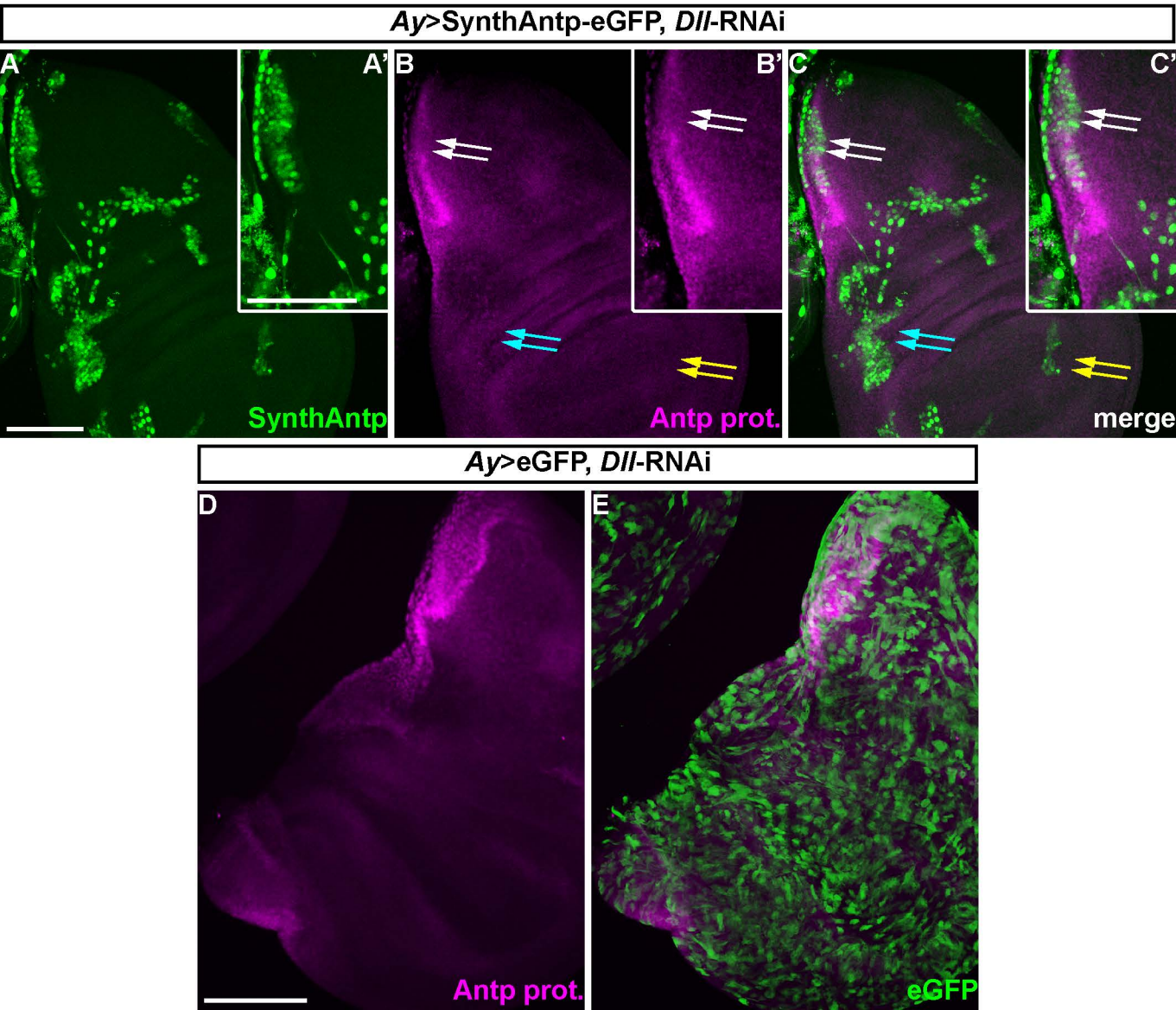


Figure 4

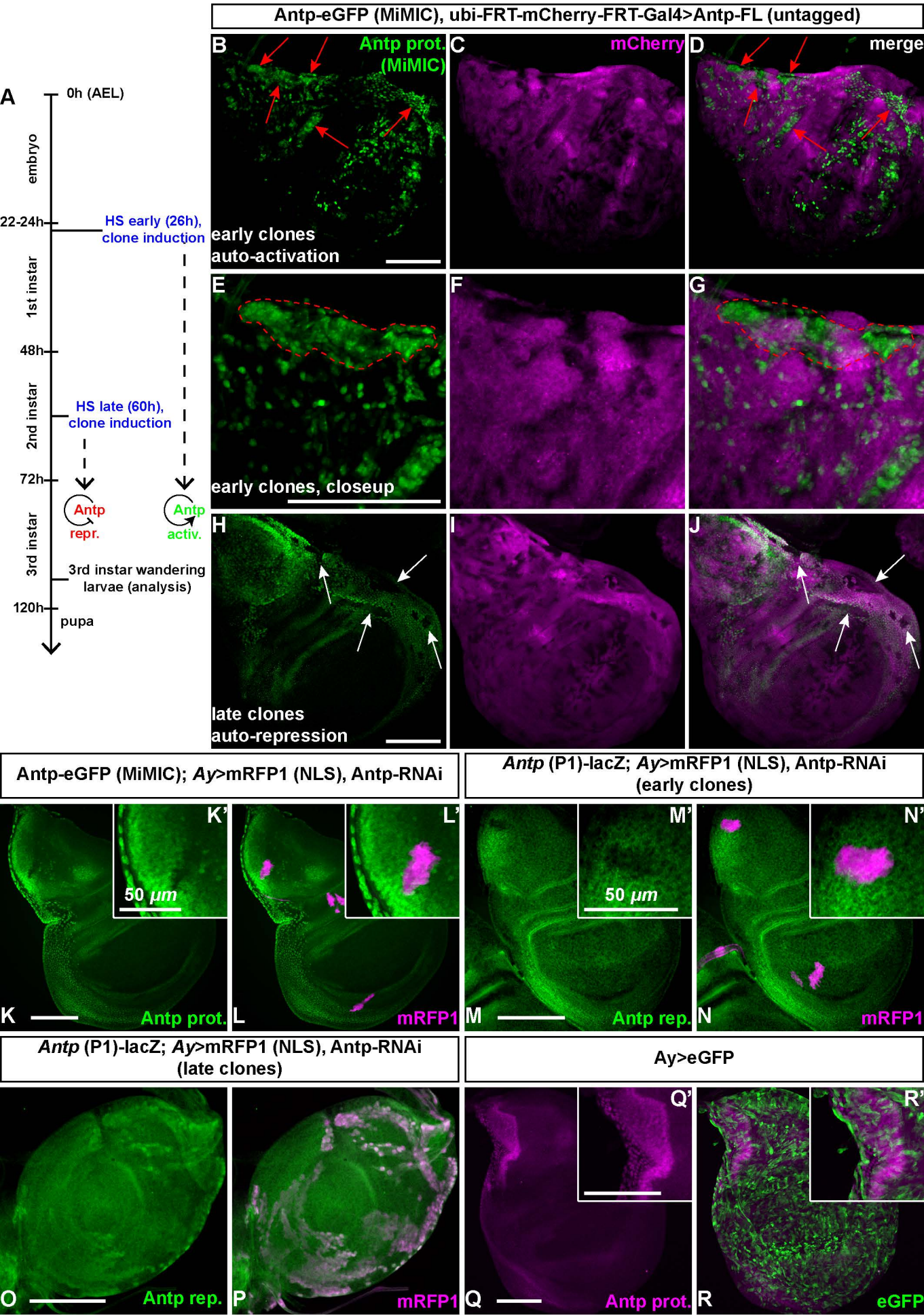

Antp-eGFP (MiMIC), ubi-FRT-mCherry-FRT-Gal4>Antp-FL (untagged)
A
0h (AEL)
embryo
22-24h
HS early (26h), clone induction
1st instar
48h
2nd instar
HS late (60h), clone induction
72h
Antp repr.
Antp activ.
3rd instar
3rd instar wandering larvae (analysis)
120h
pupa
B
Antp prot. (MiMIC)
C
mCherry
D
merge
early clones auto-activation
E
F
G
early clones, closeup
H
I
J
late clones auto-repression
Antp-eGFP (MiMIC); Ay>mRFP1 (NLS), Antp-RNAi
Antp (P1)-lacZ; Ay>mRFP1 (NLS), Antp-RNAi (early clones)
K'
L'
M'
N'
50 μm
50 μm
K
Antp prot.
L
mRFP1
M
Antp rep.
N
mRFP1
Antp (P1)-lacZ; Ay>mRFP1 (NLS), Antp-RNAi (late clones)
Ay>eGFP
Q'
R'
O
Antp rep.
P
mRFP1
Q
Antp prot.
R
eGFP

Figure 4 - figure supplement 1

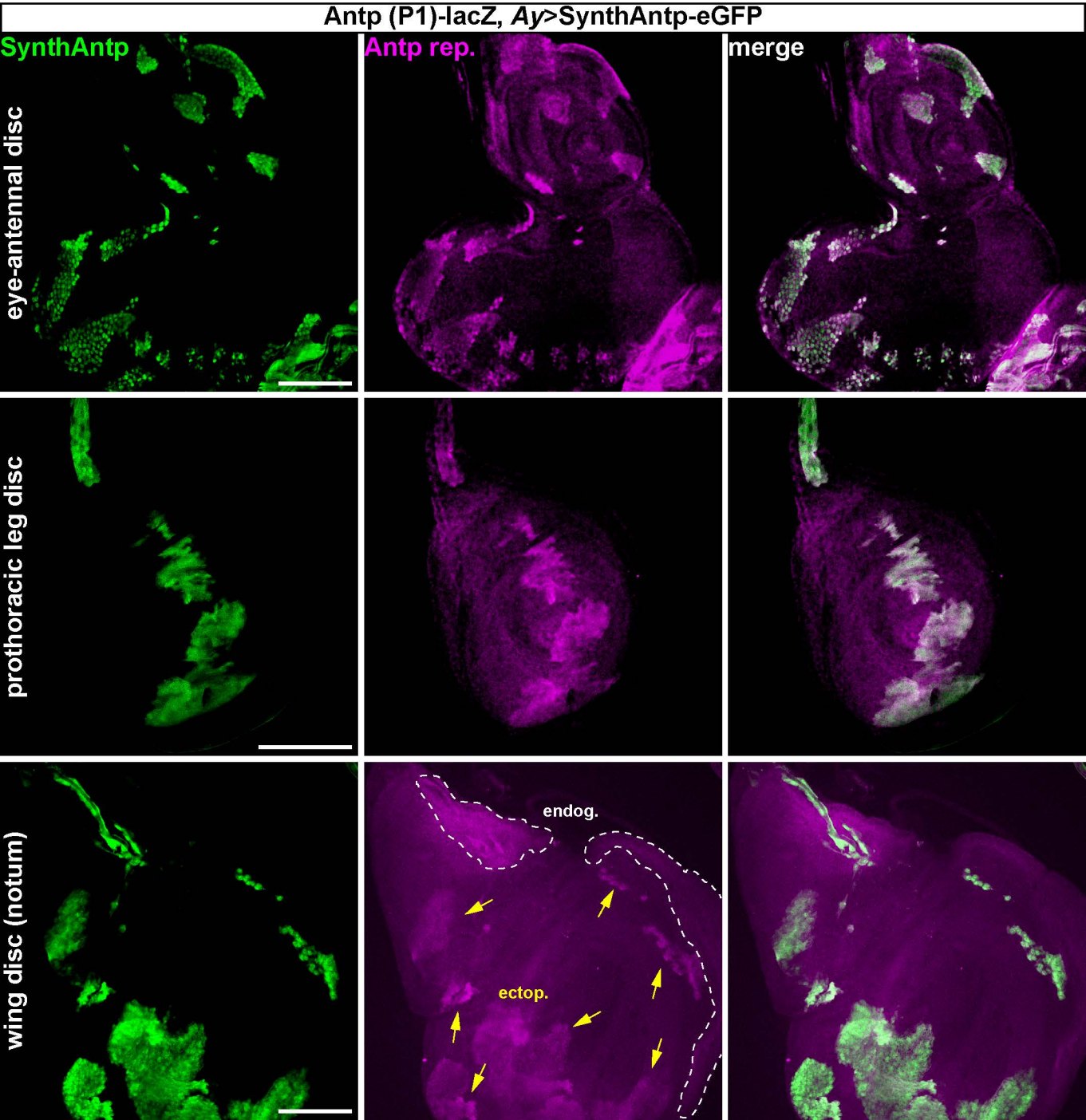


Figure 5

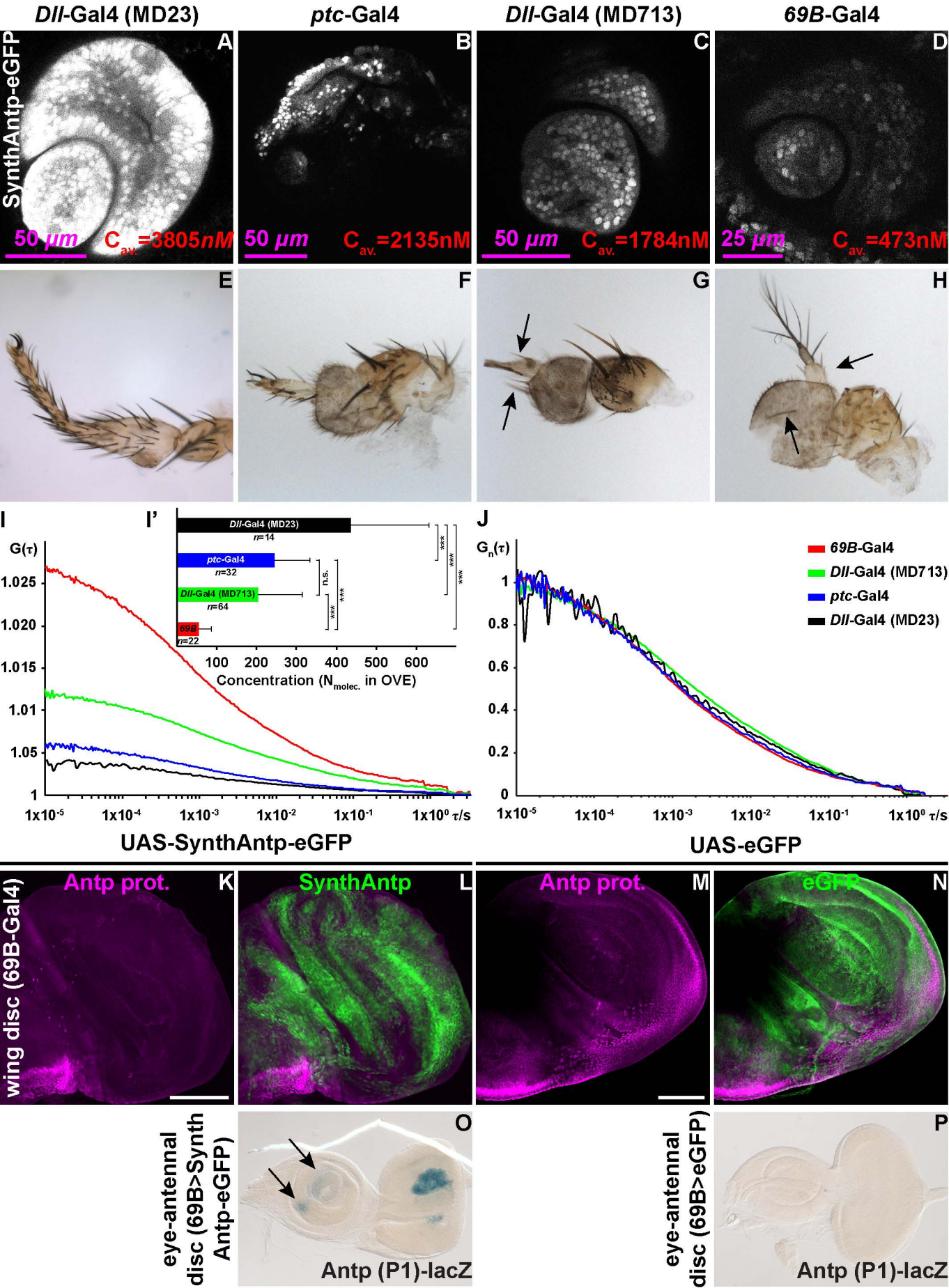

Dll-Gal4 (MD23)
ptc-Gal4
Dll-Gal4 (MD713)
69B-Gal4
SynthAntp-eGFP
A
B
C
D
50 µm
C av. =3805nM
50 µm
C av. =2135nM
50 µm
C av. =1784nM
25 µm
C av. =473nM
E
F
G
H
I
I'
G(τ)
1.025
1.020
1.015
1.01
1.05
1
1x10-5
1x10-4
1x10-3
1x10-2
1x10-1
1x100 τ/s
UAS-SynthAntp-eGFP
Dll-Gal4 (MD23)
n=14
ptc-Gal4
n=32
Dll-Gal4 (MD713)
n=64
69B
n=22
n.s.
***
100
200
300
400
500
600
Concentration (Nmolec. in OVE)
J
Gn(τ)
1
0.8
0.6
0.4
0.2
0
69B-Gal4
Dll-Gal4 (MD713)
ptc-Gal4
Dll-Gal4 (MD23)
UAS-eGFP
wing disc (69B-Gal4)
Antp prot.
K
SynthAntp
L
Antp prot.
M
eGFP
N
eye-antennal disc (69B>Synth Antp-eGFP)
O
Antp (P1)-lacZ
eye-antennal disc (69B>eGFP)
P
Antp (P1)-lacZ

Figure 5 - figure supplement 1

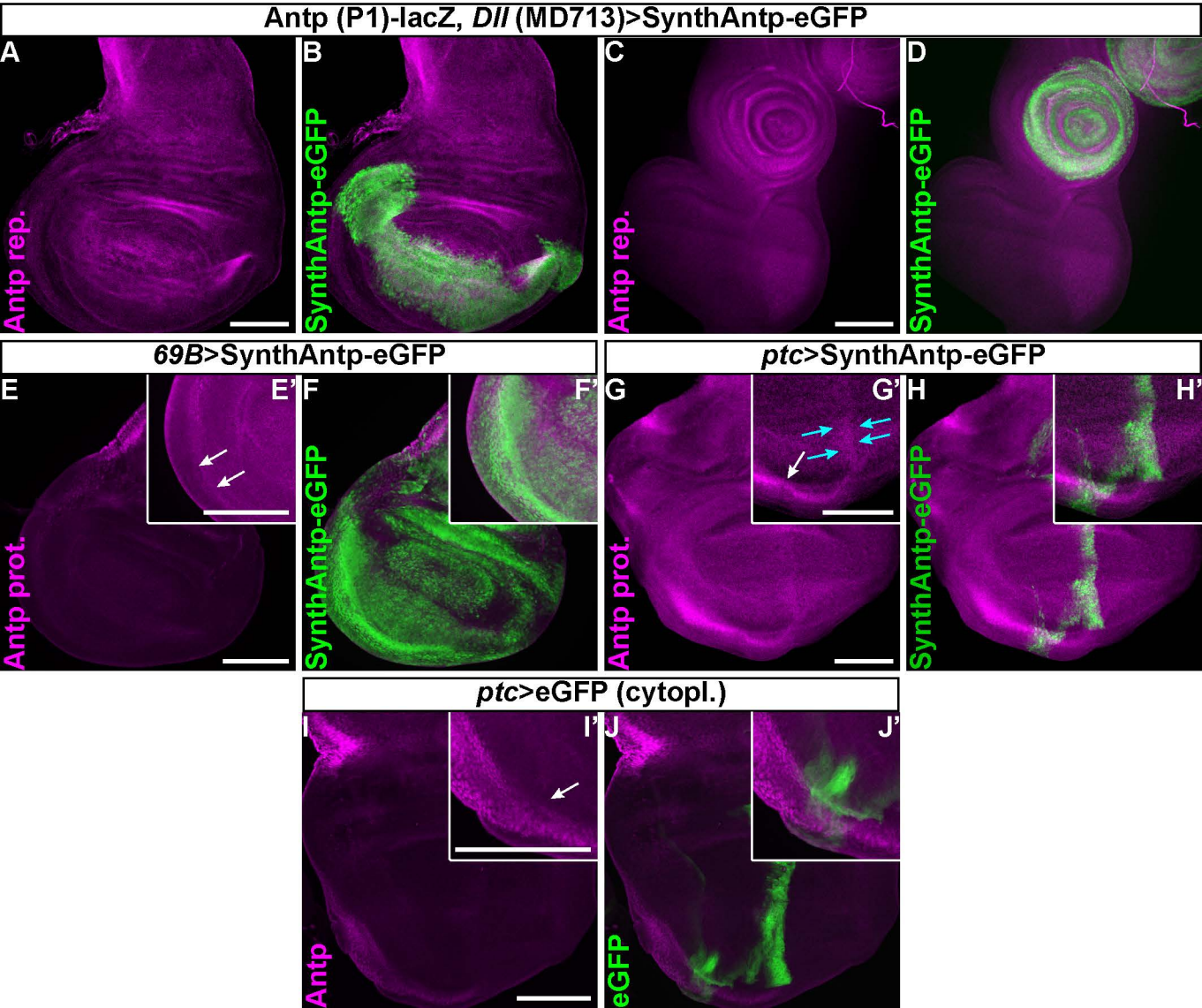


Figure 6

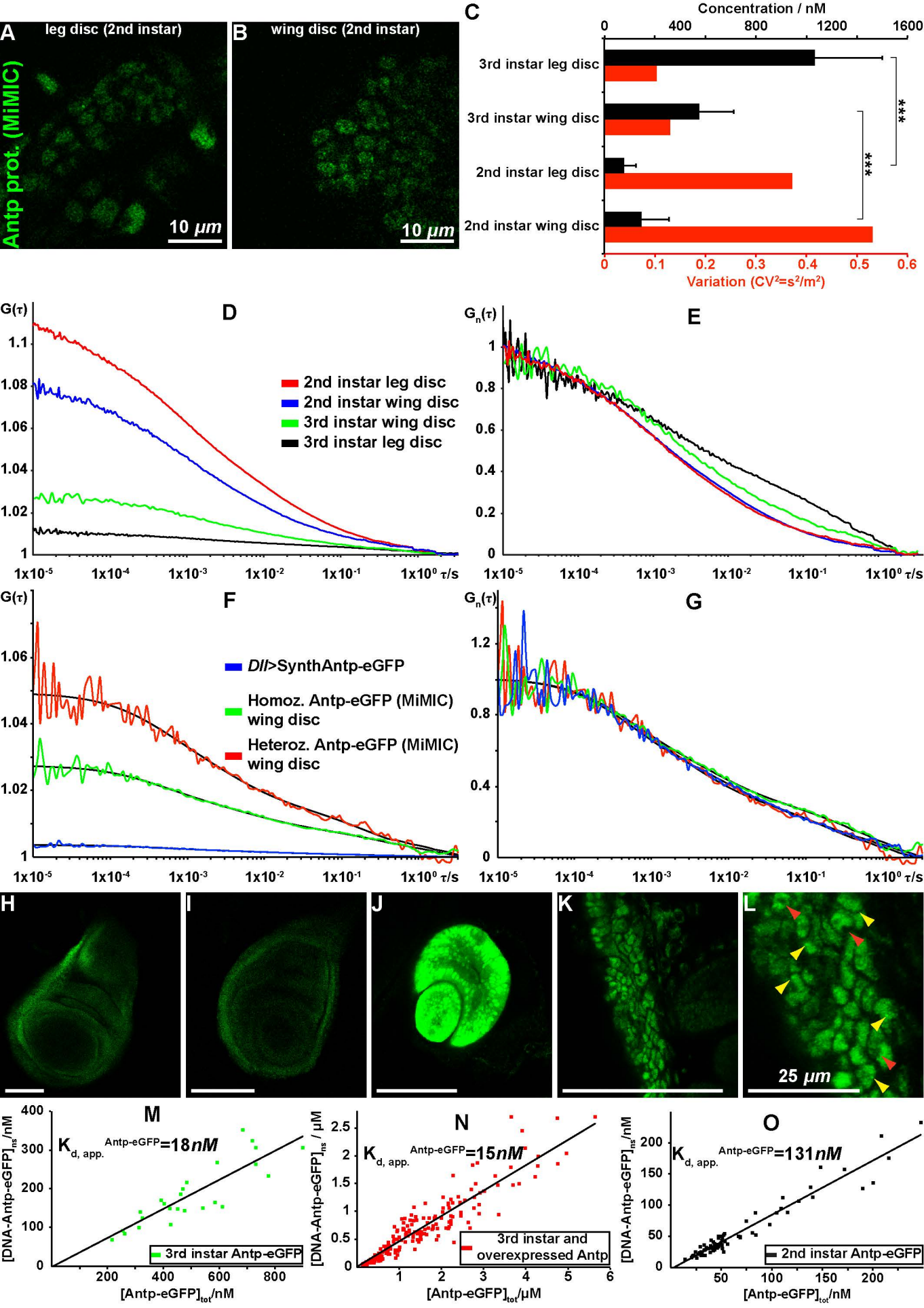

A
leg disc (2nd instar)
Antp prot. (MiMIC)
10 μm
B
wing disc (2nd instar)
10 μm
C
Concentration / nM
0
400
800
1200
1600
3rd instar leg disc
3rd instar wing disc
2nd instar leg disc
2nd instar wing disc
***
***
0
0.1
0.2
0.3
0.4
0.5
0.6
Variation (CV²=s²/m²)
D
G(τ)
2nd instar leg disc
2nd instar wing disc
3rd instar wing disc
3rd instar leg disc
1x10⁻⁵
1x10⁻⁴
1x10⁻³
1x10⁻²
1x10⁻¹
1x10⁰ τ/s
E
Gn(τ)
F
Dll>SynthAntp-eGFP
Homoz. Antp-eGFP (MiMIC) wing disc
Heteroz. Antp-eGFP (MiMIC) wing disc
G
H
I
J
K
L
25 μm
M
Kd, app. Antp-eGFP=18nM
[DNA-Antp-eGFP]ns/nM
3rd instar Antp-eGFP
[Antp-eGFP]tot/nM
N
Kd, app. Antp-eGFP=15nM
[DNA-Antp-eGFP]ns / μM
3rd instar and overexpressed Antp
[Antp-eGFP]tot/μM
O
Kd, app. Antp-eGFP=131nM
[DNA-Antp-eGFP]ns/nM
2nd instar Antp-eGFP
[Antp-eGFP]tot/nM

Figure 6 - figure supplement 1

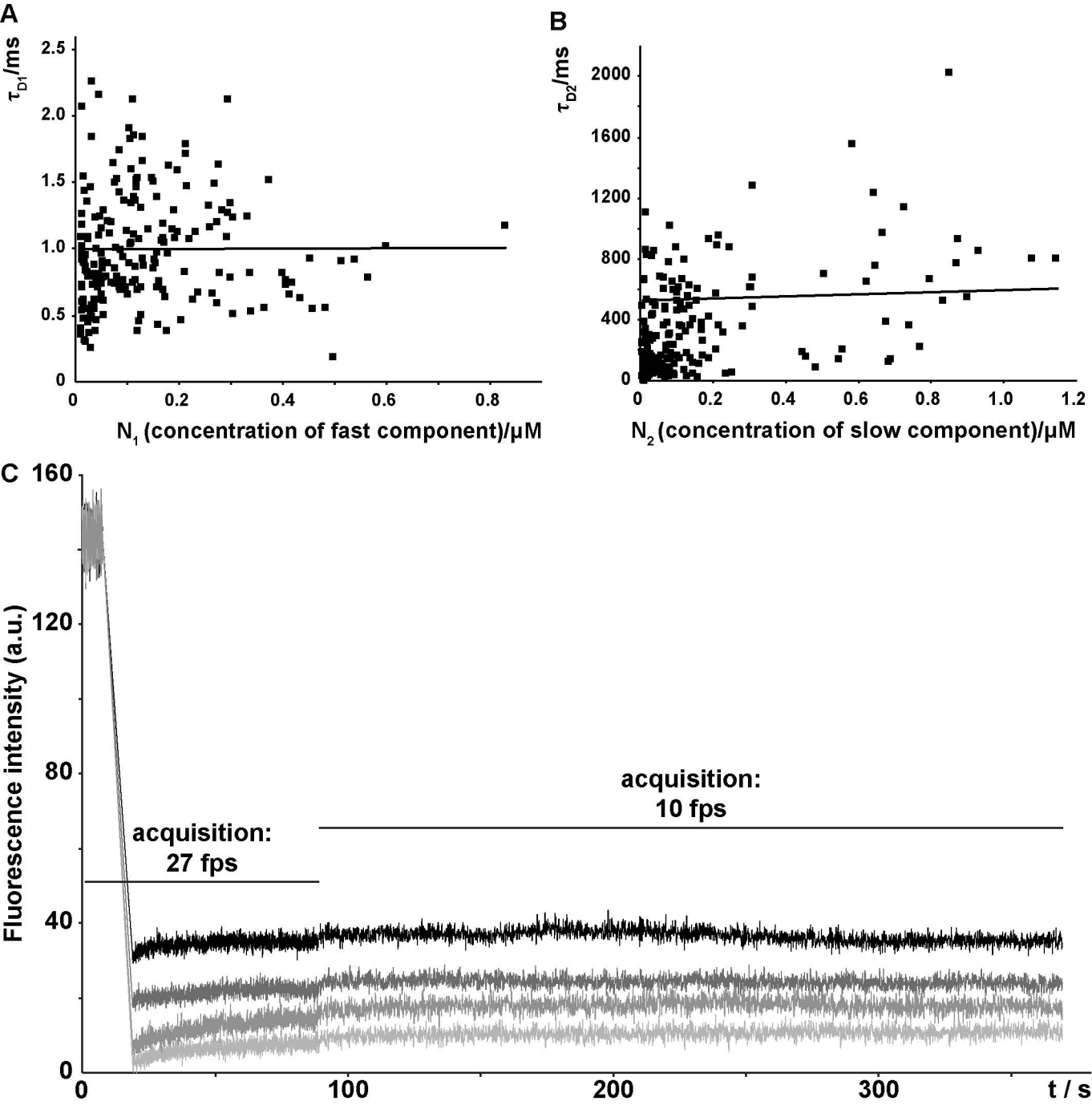


Figure 6 - figure supplement 2

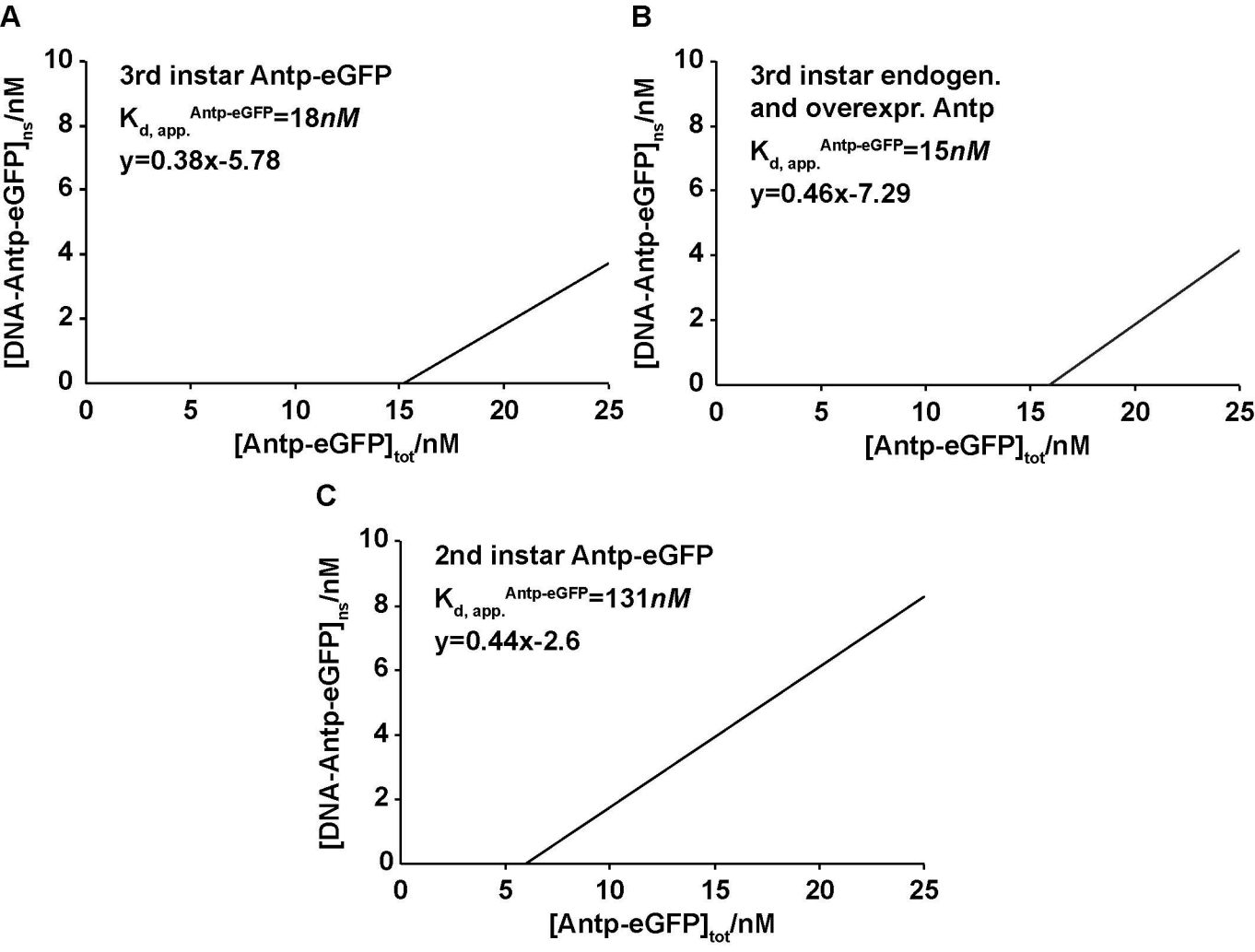


Figure 7

A

WT

a) ...GMPSPLYPWMRSQFGKCQERKRGR...

N-term. long linker hexapeptide homeodomain

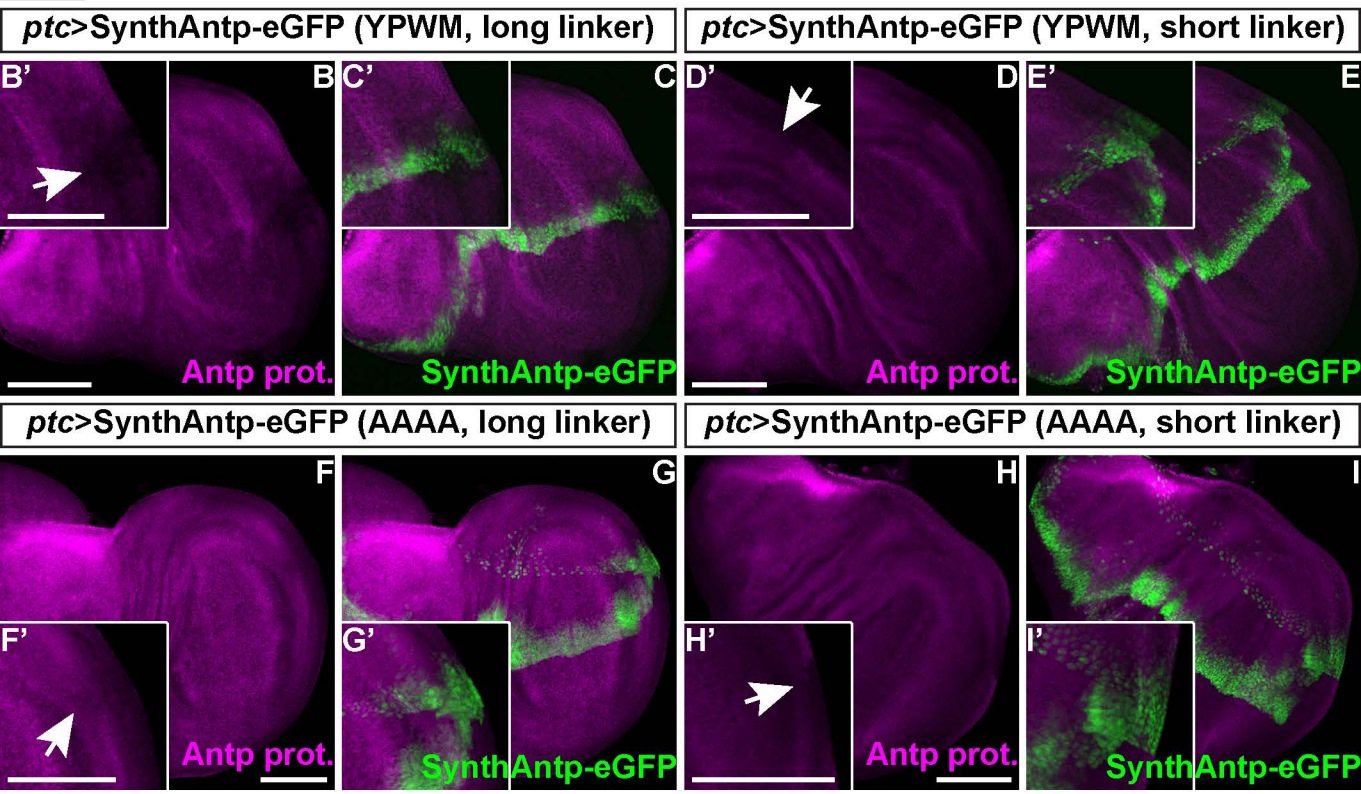
b) ...GMPSPLYPWMRSQFERKRGR...

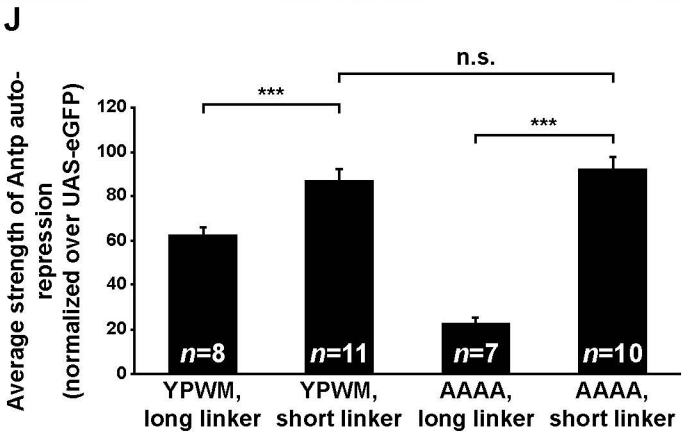
short linker

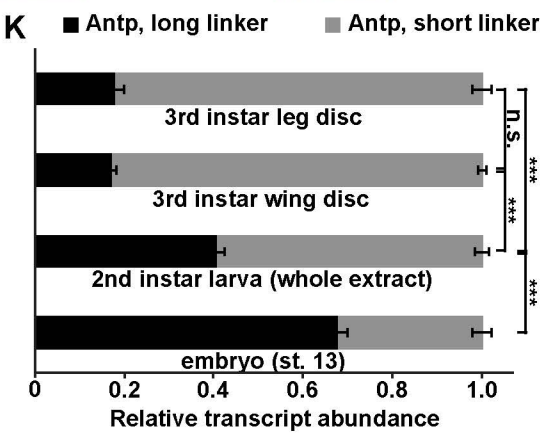
YPWM> AAAA

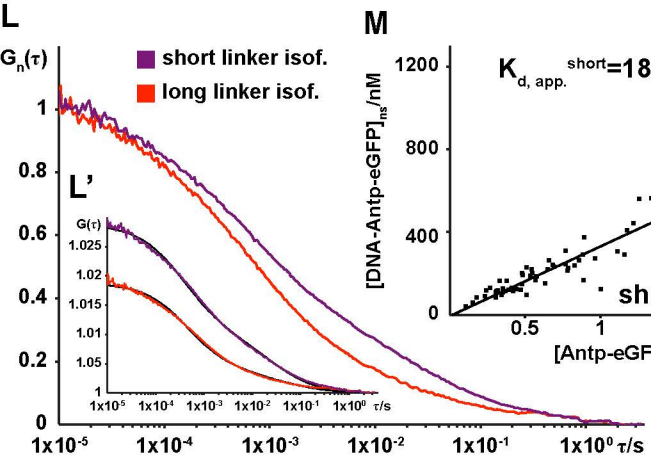
c) ...GMPSPLAAAARSQFGKCQERKRGR...

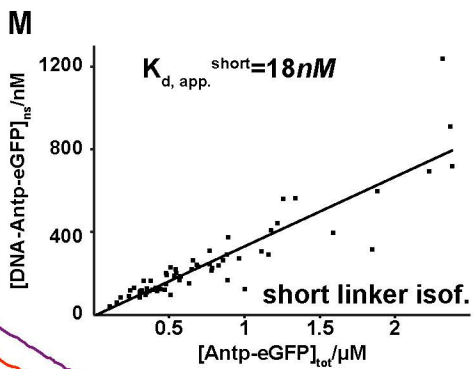
d) ...GMPSPLAAAARSQFERKRGR...











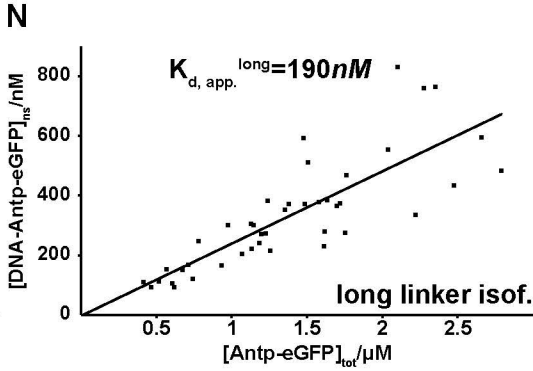


Figure 7 - figure supplement 1

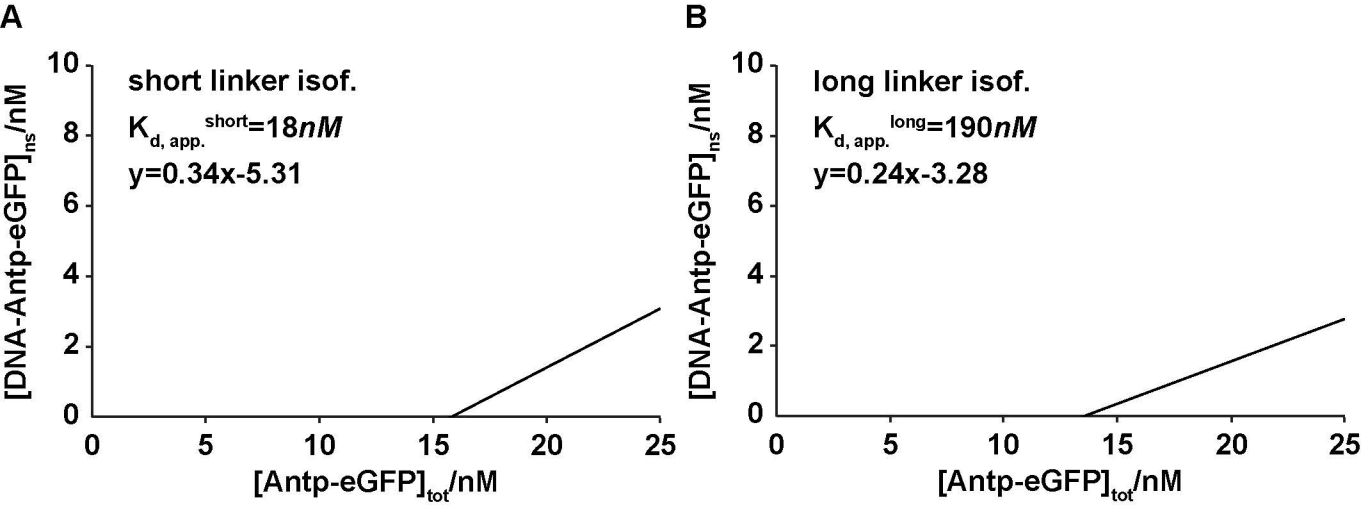


Figure 7 - figure supplement 2

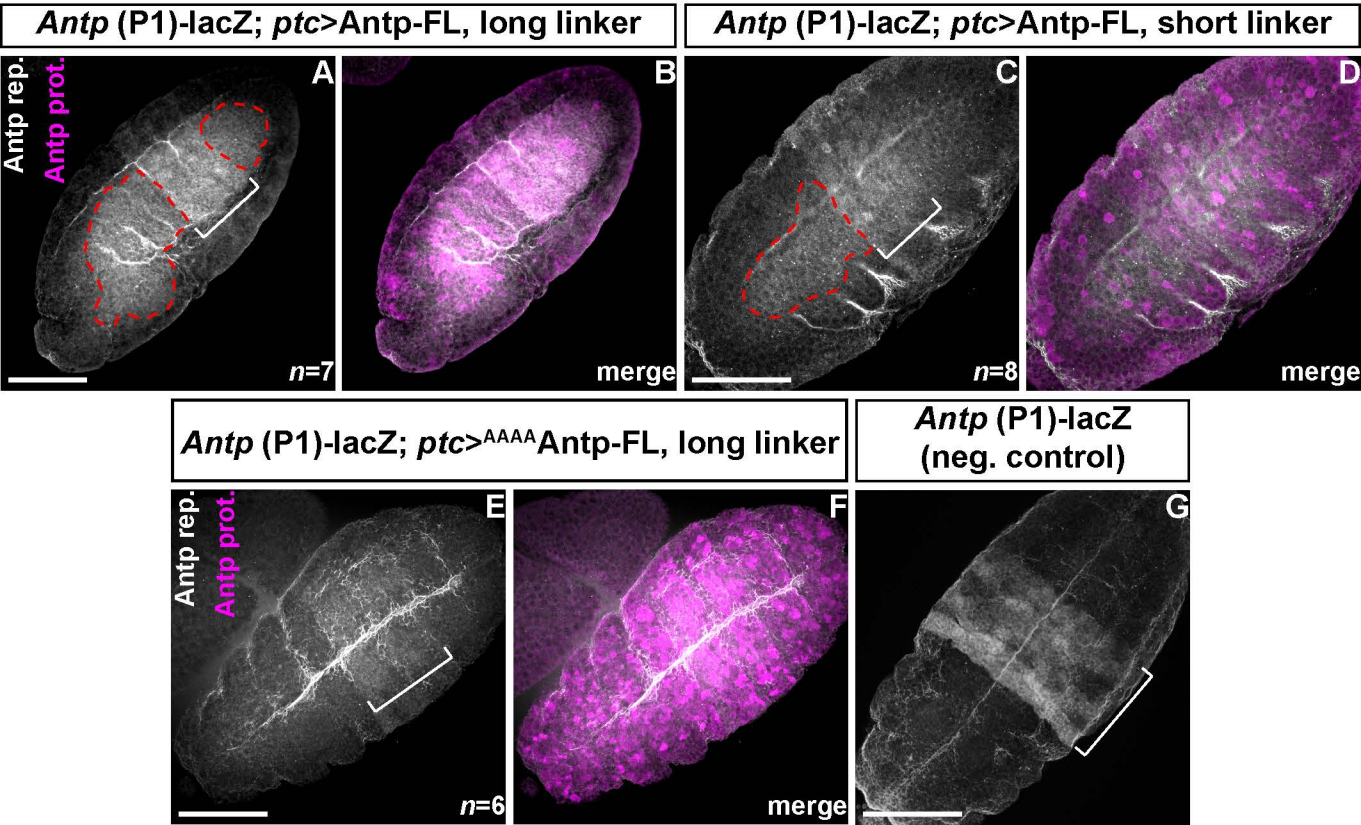
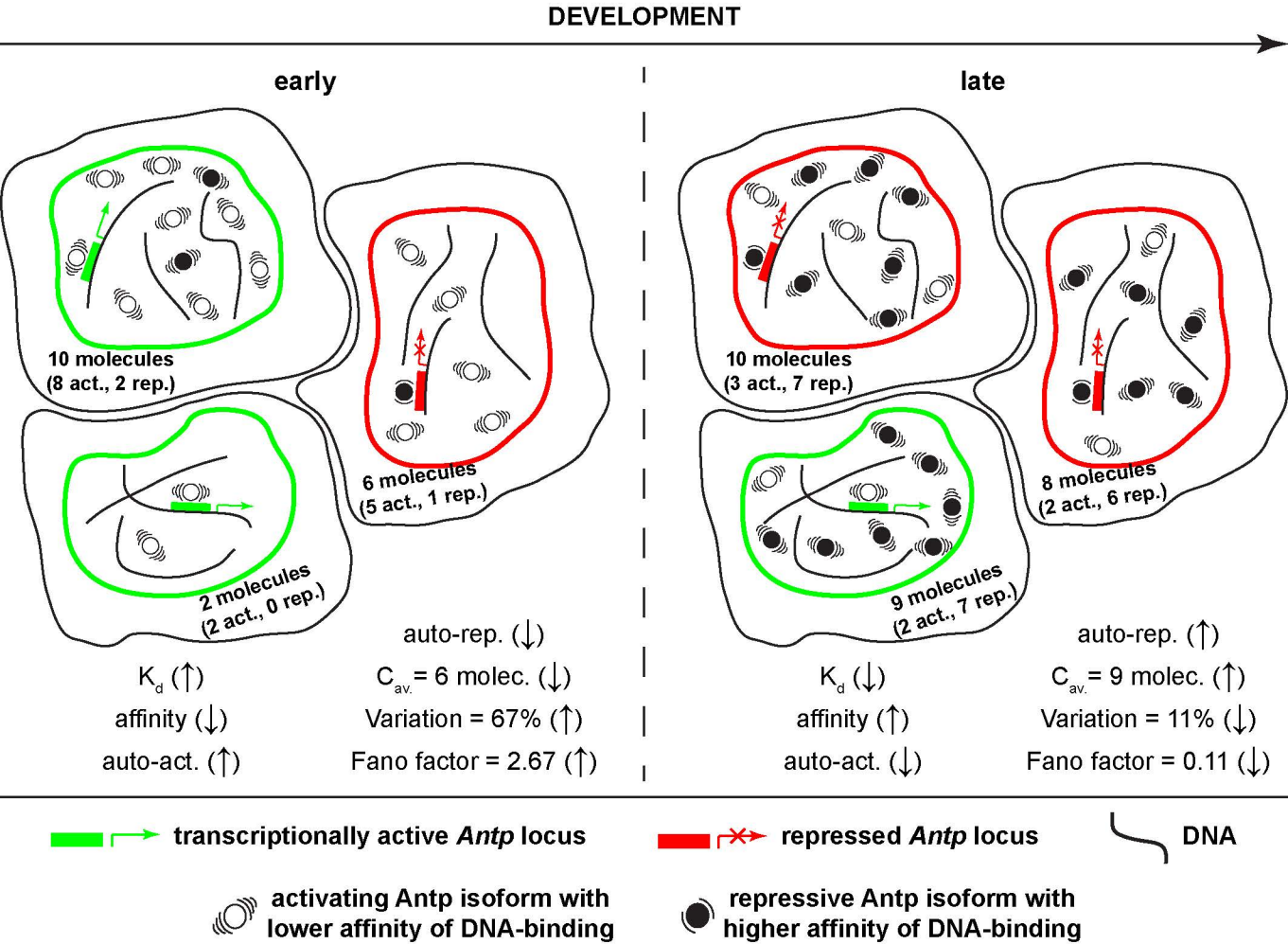


Figure 8


DEVELOPMENT
early
late
10 molecules
(8 act., 2 rep.)
6 molecules
(5 act., 1 rep.)
2 molecules
(2 act., 0 rep.)
10 molecules
(3 act., 7 rep.)
8 molecules
(2 act., 6 rep.)
9 molecules
(2 act., 7 rep.)
auto-rep. (↓)
K_d (↑)
$C_{av.}$ = 6 molec. (↓)
affinity (↓)
Variation = 67% (↑)
auto-act. (↑)
Fano factor = 2.67 (↑)
auto-rep. (↑)
K_d (↓)
$C_{av.}$ = 9 molec. (↑)
affinity (↑)
Variation = 11% (↓)
auto-act. (↓)
Fano factor = 0.11 (↓)
transcriptionally active *Antp* locus
repressed *Antp* locus
DNA
activating Antp isoform with lower affinity of DNA-binding
repressive Antp isoform with higher affinity of DNA-binding

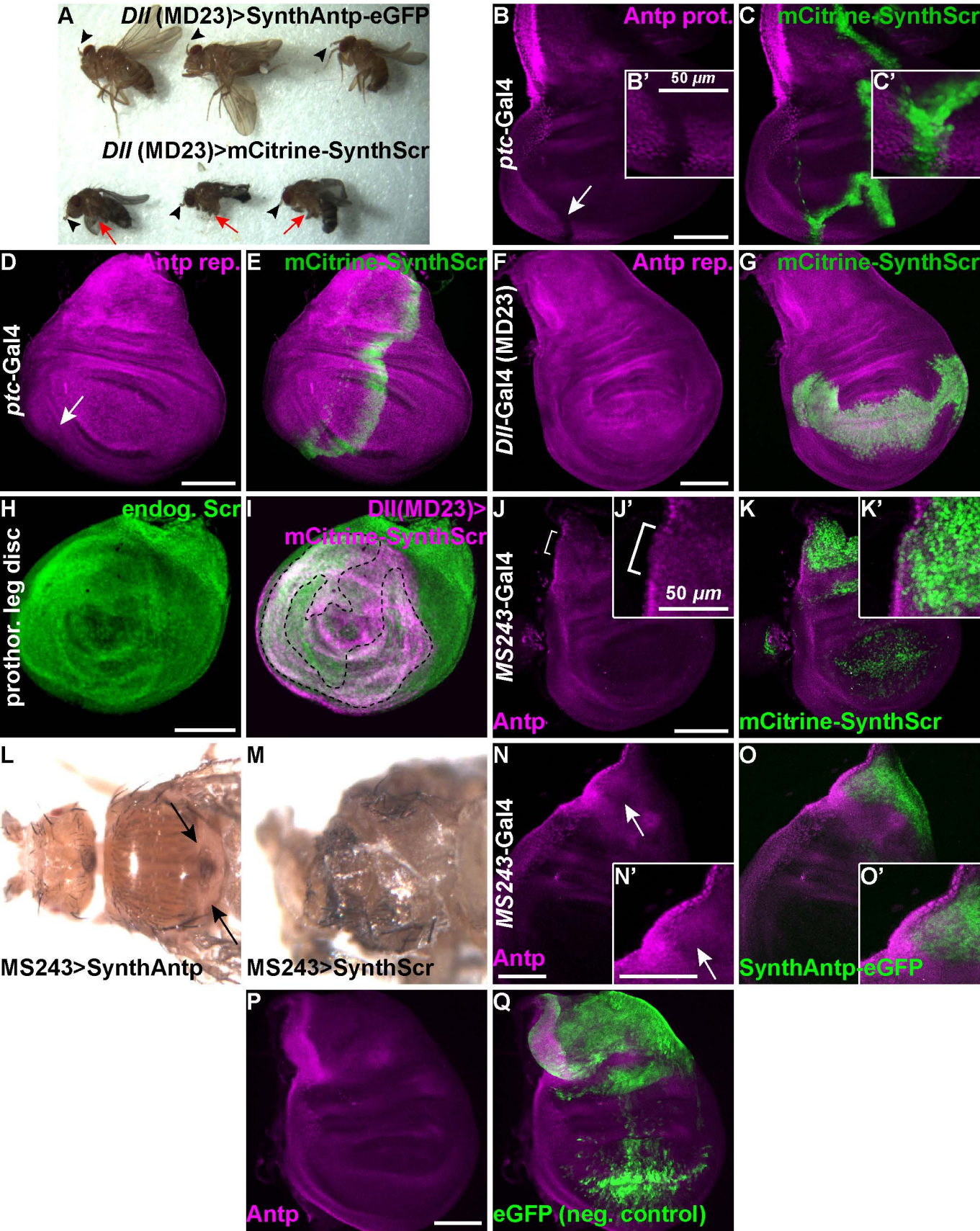



Figure 8 - figure supplement 1